

WO0021985

Publication Title:

GENES ENCODING OLFACTORY RECEPTORS AND BIALLELIC MARKERS THEREOF

Abstract:

Abstract of WO0021985

The invention concerns the genomic sequence and coding regions of a new olfactory receptor gene cluster. The invention also concerns polypeptides encoded by the olfactory receptor genes as well as to methods and kits for detecting these polynucleotides and screening substances interacting with these polypeptides. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the olfactory receptor gene useful in genetic analysis. Data supplied from the esp@cenet database - Worldwide

Courtesy of <http://v3.espacenet.com>



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C07K 14/00		A2	(11) International Publication Number: WO 00/21985 (43) International Publication Date: 20 April 2000 (20.04.00)
<p>(21) International Application Number: PCT/IB99/01729</p> <p>(22) International Filing Date: 13 October 1999 (13.10.99)</p> <p>(30) Priority Data: 60/104,299 14 October 1998 (14.10.98) US</p> <p>(71) Applicant (for all designated States except US): GENSET [FR/FR]; 24, rue Royale, F-75008 Paris (FR).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): BOUGUELERET, Lydie [FR/FR]; 108, avenue Victor Hugo, F-92170 Vanves (FR). MALEKZADEH, Kattayoun [IR/FR]; 41, rue des Laitieres, F-94300 Vincennes (FR).</p> <p>(74) Common Representative: GENSET; 24, rue Royale, F-75008 Paris (FR).</p>		<p>(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).</p> <p>Published Without international search report and to be republished upon receipt of that report.</p>	
<p>(54) Title: GENES ENCODING OLFACTORY RECEPTORS AND BIALLELIC MARKERS THEREOF</p> <p>(57) Abstract</p> <p>The invention concerns the genomic sequence and coding regions of a new olfactory receptor gene cluster. The invention also concerns polypeptides encoded by the olfactory receptor genes as well as to methods and kits for detecting these polynucleotides and screening substances interacting with these polypeptides. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the olfactory receptor gene useful in genetic analysis.</p>			

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

GENES ENCODING OLFACTORY RECEPTORS AND BIALLELIC MARKERS THEREOF

FIELD OF THE INVENTION

The present invention pertains to a purified or isolated nucleic acid comprising ten open reading Frames (ORFs) encoding ten different olfactory receptor-like proteins, non-coding regions flanking the ORFs as well as fragments thereof. The invention also provides recombinant expression vectors and recombinant cell hosts containing a nucleic acid encoding said olfactory receptor proteins. The invention also concerns the olfactory receptor proteins encoded by these ORFs as well as polypeptides that are homologous to said olfactory receptor proteins and the peptide fragments of both the olfactory receptor proteins and their homologous polypeptide counterparts. The invention also deals with antibodies directed specifically against such polypeptides that are useful as diagnostic reagents. The invention further encompasses biallelic markers of the olfactory receptor gene useful in genetic analysis. The invention also deals with methods and kits for the detection of the olfactory receptor proteins and with methods and kits for screening ligand molecules binding to these proteins.

BACKGROUND OF THE INVENTION

Throughout this application, various bibliographic publications are cited. Full bibliographic references for these publications may be found at the end of this application, preceding the sequence listing and the claims.

20 OLFACTORY SYSTEM

The olfactory receptor cells, the first cells in the pathway that give rise to the sense of smell, lie in a small patch of membrane, the olfactory epithelium, in the upper part of the nasal cavity. These cells are specialized afferent neurons that have an enlarged extension analogous to a dendrite. Several long hairlike processes extend out from this extension along the surface of the olfactory epithelium where they are bathed in mucus. The hairlike processes contain the receptor proteins for olfactory stimuli. The axons of these neurons form the olfactory nerve.

For the detection of an odorous substance which is called an odorant, molecules of the substance must first diffuse into the air and pass into the nose to the region of the olfactory epithelium. Once there, they dissolve in the mucus that covers the epithelium and then bind to specific receptor proteins on the cilia.

Although there are many thousands of olfactory neurons, each contains one, or at most a few, of the 1,000 or so different receptor types, each of which responds only to a specific chemically related group of odorant molecules. Each odorant has characteristic chemical groups that distinguish it from other odorants, and each of these groups activates a different receptor type. Thus the identity

of a particular odorant is determined by the activation of a precise combination of receptors, each of which is contained in a distinct group of olfactory neurons.

The axons of the olfactory neurons synapse in the brain structures known as olfactory bulbs, which lie on the undersurface of the frontal lobes. Axons from olfactory neurons sharing a common 5 receptor specificity synapse together on certain olfactory-bulb neurons, thereby maintaining the specificity of the original stimuli.

OLFACtORY RECEPTORS

In contrast with the immunoglobulin system, the diversity of olfactory receptors is encoded by a large germ-line repertoire of olfactory receptor genes. The size of the olfactory receptor gene 10 family in the human genome is unknown but it has been estimated to encompass 200 to 1,000 genes.

The locations of only a few human genes have been determined to date. The picture that has emerged so far is that several large clusters of olfactory genes and pseudogenes span hundreds of kilobases on several chromosomes. Using FISH analyses, more than 25 distinct locations of olfactory receptors gene have been identified in the human genome.

15 In mammals, the olfactory epithelium appears to be organized into distinct topographic regions or zones in which expression of a particular receptor gene appears to be restricted to one of the four zones in the epithelium. Within the zone, the distribution of neurons expressing a given receptor is random. Chromosomal mapping studies have revealed clusters of odorant receptor genes at a single locus, and numerous such loci have been mapped to different chromosomes. However, 20 receptors expressed in the same zone map to different loci, and a single locus can contain genes expressed in different zones. A putative odorant receptor promoter, consisting of the 6.7 kb DNA fragment upstream of the receptor coding region, has been shown to be sufficient to direct olfactory receptor expression in a tissue-specific, zonal-specific manner.

Olfactory receptors share a seven-transmembrane domain structure (TM1 to TM7) with 25 many neurotransmitter and hormone receptors. They show a high degree of sequence similarity in some conserved domains (TM2 and TM7) as well as regions of diversity (TM3, TM4, TM5, and TM6). They are responsible for the recognition and G protein-mediated transduction of odorant signals. The genes encoding these receptors are devoid of introns within their coding regions.

Olfactory receptors display all hallmarks of the G-protein coupled receptor superfamily but 30 have also some unique motifs. Most notably they appear to be minimal in structure with very short cytoplasmic and extracellular loops. In addition, they display a striking structural diversity in the third, fourth and fifth transmembrane domains which are supposed to form the hydrophobic core of these proteins, and may form the ligand binding site of the receptors.

An understanding of the genetic basis of olfaction and a knowledge of olfactory receptors 35 are important to enable the design of fragrance, the identification of compounds which control appetite, or the detection of compounds which can be harmful or dangerous.

SUMMARY OF THE INVENTION

This invention provides a nucleic acid molecule encoding ten different olfactory receptor-like proteins (OLF).

The invention also deals with a nucleic acid molecule comprising a nucleotide sequence 5 encoding an olfactory receptor-like protein, which nucleotide sequence is selected from the group consisting of SEQ ID Nos 2-11, as well as with the corresponding polypeptide encoded by this nucleotide sequence and with antibodies directed against the corresponding polypeptide.

Oligonucleotide probes or primers hybridizing specifically with an olfactory receptor genomic sequence are also part of the present invention, as well as DNA amplification and detection 10 methods using said primers and probes.

The invention also concerns a purified and/or isolated biallelic marker located in the sequence of the olfactory receptor gene cluster of the invention, wherein said biallelic marker is useful as a diagnostic tool in order to detect an allele associated with a specific phenotype as regards to the olfaction system, including an alteration of the olfactory perception of substances or 15 molecules.

A further object of the invention consists of recombinant vectors comprising any of the nucleic acid sequences described above, and in particular of recombinant vectors comprising a sequence encoding an olfactory receptor protein, as well as of cell hosts and transgenic non human animals comprising said nucleic acid sequences or recombinant vectors.

20 A further object of the invention consists of methods for screening substances or molecules interacting with an olfactory receptor encoded by any of the nucleic acid molecule described above.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 : Alignment of the amino acid sequences of the olfactory polypeptides encoded by 25 the Open Reading Frames of the olfactory receptor gene cluster of the invention. The lower line represents the consensus sequence. The locations of the seven transmembrane domains TM1 to TM7 are boxed.

BRIEF DESCRIPTION OF THE SEQUENCES PROVIDED IN THE SEQUENCE LISTING

30 SEQ ID No 1 contains the olfactory receptor genomic sequence.

SEQ ID Nos 2-11 contains the nucleotide sequences of the open reading frame sequences of SEQ ID No 1 encoding the OLF1 to OLF10 polypeptides.

SEQ ID No 12-21 contain the amino acid sequence of OLF1 to OLF10 polypeptides encoded by the open reading frames of SEQ ID Nos 2-11.

SEQ ID Nos 22-25 contain the amplification primers used for FISH experiments described in Example 1.

SEQ ID No 26 contains a primer containing the additional PU 5' sequence described further in Example 3.

5 SEQ ID No 27 contains a primer containing the additional RP 5' sequence described further in Example 3.

In accordance with the regulations relating to Sequence Listings, the following codes have been used in the Sequence Listing to indicate the locations of biallelic markers within the sequences and to identify each of the alleles present at the polymorphic base. The code "r" in the sequences 10 indicates that one allele of the polymorphic base is a guanine, while the other allele is an adenine. The code "y" in the sequences indicates that one allele of the polymorphic base is a thymine, while the other allele is a cytosine. The code "m" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an cytosine. The code "k" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a thymine. 15 The code "s" in the sequences indicates that one allele of the polymorphic base is a guanine, while the other allele is a cytosine. The code "w" in the sequences indicates that one allele of the polymorphic base is an adenine, while the other allele is an thymine.

The nucleotide code of the original allele for each biallelic marker is the following:

	Biallelic marker	Original allele
20	99-13670-305	G
	99-13669-471	G
	99-13666-275	A
	99-13664-221	T
	99-13663-218	G
25	99-13660-277	C
	99-13652-407	G
	99-13652-357	A
	99-13652-308	A
	99-13671-396	A
30	99-13649-286	C
	99-13648-259	G
	99-13647-278	G

DETAILED DESCRIPTION OF THE INVENTION

35 The aim of the present invention is to provide polynucleotides and polypeptides related to novel olfactory receptors, notably useful in order to design suitable means for detecting specific odorant molecules in a material sample, particularly in a material sample suspected to contain an odorant molecule that consists of one of the specific ligands for the olfactory receptors of the invention.

DEFINITIONS

Before describing the invention in greater detail, the following definitions are set forth to illustrate and define the meaning and scope of the terms used to describe the invention herein.

General definitions

5 The terms "olfactory receptor gene" or "OLF1 to OLF10" genes, when used herein, encompasses genomic, mRNA and cDNA sequences encoding the OLF1 to OLF10 olfactory receptor proteins.

The term "heterologous protein", when used herein, is intended to designate any protein or polypeptide other than the OLF1 to OLF10 proteins.

10 The term "isolated" requires that the material be removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polynucleotide or polypeptide present in a living animal is not isolated, but the same polynucleotide or DNA or polypeptide, separated from some or all of the coexisting materials in the natural system, is isolated. Such polynucleotide could be part of a vector and/or such polynucleotide or polypeptide 15 could be part of a composition, and still be isolated in that the vector or composition is not part of its natural environment.

The term "purified" does not require absolute purity; rather, it is intended as a relative definition. Purification of starting material or natural material to at least one order of magnitude, preferably two or three orders, and more preferably four or five orders of magnitude is expressly 20 contemplated. As an example, purification from 0.1 % concentration to 10 % concentration is two orders of magnitude. The term "purified polynucleotide" is used herein to describe a polynucleotide or polynucleotide vector of the invention which has been separated from other compounds including, but not limited to other nucleic acids, carbohydrates, lipids and proteins (such as the enzymes used in the synthesis of the polynucleotide), or the separation of covalently closed polynucleotides from 25 linear polynucleotides. A polynucleotide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polynucleotide sequence and conformation (linear versus covalently close). A substantially pure polynucleotide typically comprises about 50%, preferably 60 to 90% weight/weight of a nucleic acid sample, more usually about 95%, and preferably is over about 99% pure. Polynucleotide purity or homogeneity is indicated by a number of means well 30 known in the art, such as agarose or polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polynucleotide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

The term "polypeptide" refers to a polymer of amino acids without regard to the length of the polymer; thus, peptides, oligopeptides, and proteins are included within the definition of 35 polypeptide. This term also does not specify or exclude post-expression modifications of polypeptides, for example, polypeptides which include the covalent attachment of glycosyl groups, acetyl groups, phosphate groups, lipid groups and the like are expressly encompassed by the term

polypeptide. Also included within the definition are polypeptides which contain one or more analogs of an amino acid (including, for example, non-naturally occurring amino acids, amino acids which only occur naturally in an unrelated biological system, modified amino acids from mammalian systems etc.), polypeptides with substituted linkages, as well as other modifications

5 known in the art, both naturally occurring and non-naturally occurring.

The term "recombinant polypeptide" is used herein to refer to polypeptides that have been artificially designed and which comprise at least two polypeptide sequences that are not found as contiguous polypeptide sequences in their initial natural environment, or to refer to polypeptides which have been expressed from a recombinant polynucleotide.

10 The term "purified polypeptide" is used herein to describe a polypeptide of the invention which has been separated from other compounds including, but not limited to nucleic acids, lipids, carbohydrates and other proteins. A polypeptide is substantially pure when at least about 50%, preferably 60 to 75% of a sample exhibits a single polypeptide sequence. A substantially pure polypeptide typically comprises about 50%, preferably 60 to 90% weight/weight of a protein sample, 15 more usually about 95%, and preferably is over about 99% pure. Polypeptide purity or homogeneity is indicated by a number of means well known in the art, such as polyacrylamide gel electrophoresis of a sample, followed by visualizing a single polypeptide band upon staining the gel. For certain purposes higher resolution can be provided by using HPLC or other means well known in the art.

20 As used herein, the term "non-human animal" refers to any non-human vertebrate, birds and more usually mammals, preferably primates, farm animals such as swine, goats, sheep, donkeys, and horses, rabbits or rodents, more preferably rats or mice. As used herein, the term "animal" is used to refer to any vertebrate, preferable a mammal. Both the terms "animal" and "mammal" expressly embrace human subjects unless preceded with the term "non-human".

25 As used herein, the term "antibody" refers to a polypeptide or group of polypeptides which are comprised of at least one binding domain, where an antibody binding domain is formed from the folding of variable domains of an antibody molecule to form three-dimensional binding spaces with an internal surface shape and charge distribution complementary to the features of an antigenic determinant of an antigen, which allows an immunological reaction with the antigen. Antibodies include recombinant proteins comprising the binding domains, as wells as fragments, including Fab, 30 Fab', F(ab)₂, and F(ab')₂ fragments.

As used herein, an "antigenic determinant" is the portion of an antigen molecule, in this case a OLF1 to OLF10 polypeptide, that determines the specificity of the antigen-antibody reaction. An "epitope" refers to an antigenic determinant of a polypeptide. An epitope can comprise as few as 3 amino acids in a spatial conformation which is unique to the epitope. Generally an epitope 35 comprises at least 6 such amino acids, and more usually at least 8-10 such amino acids. Methods for determining the amino acids which make up an epitope include x-ray crystallography, 2-dimensional

nuclear magnetic resonance, and epitope mapping e.g. the Pepscan method described by Geysen et al. 1984; PCT Publication No. WO 84/03564; and PCT Publication No. WO 84/03506.

Throughout the present specification, the expression "nucleotide sequence" may be employed to designate indifferently a polynucleotide or a nucleic acid. More precisely, the 5 expression "nucleotide sequence" encompasses the nucleic material itself and is thus not restricted to the sequence information (i.e. the succession of letters chosen among the four base letters) that biochemically characterizes a specific DNA or RNA molecule.

As used interchangeably herein, the terms "nucleic acids", "oligonucleotides", and "polynucleotides" include RNA, DNA, or RNA/DNA hybrid sequences of more than one nucleotide 10 in either single chain or duplex form. The term "nucleotide" as used herein as an adjective to describe molecules comprising RNA, DNA, or RNA/DNA hybrid sequences of any length in single-stranded or duplex form. The term "nucleotide" is also used herein as a noun to refer to individual nucleotides or varieties of nucleotides, meaning a molecule, or individual unit in a larger nucleic acid molecule, comprising a purine or pyrimidine, a ribose or deoxyribose sugar moiety, and a 15 phosphate group, or phosphodiester linkage in the case of nucleotides within an oligonucleotide or polynucleotide. The term "nucleotide" is also used herein to encompass "modified nucleotides" which comprise at least one modifications (a) an alternative linking group, (b) an analogous form of purine, (c) an analogous form of pyrimidine, or (d) an analogous sugar, for examples of analogous linking groups, purine, pyrimidines, and sugars see for example PCT publication No. WO 95/04064. 20 The polynucleotide sequences of the invention may be prepared by any known method, including synthetic, recombinant, *ex vivo* generation, or a combination thereof, as well as utilizing any purification methods known in the art.

A "promoter" refers to a DNA sequence recognized by the synthetic machinery of the cell required to initiate the specific transcription of a gene.

25 A sequence which is "operably linked" to a regulatory sequence such as a promoter means that said regulatory element is in the correct location and orientation in relation to the nucleic acid to control RNA polymerase initiation and expression of the nucleic acid of interest. As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the 30 transcription of the coding sequence. More precisely, two DNA molecules (such as a polynucleotide containing a promoter region and a polynucleotide encoding a desired polypeptide or polynucleotide) are said to be "operably linked" if the nature of the linkage between the two polynucleotides does not (1) result in the introduction of a frame-shift mutation or (2) interfere with the ability of the polynucleotide containing the promoter to direct the transcription of the coding 35 polynucleotide.

The term "vector" is used herein to designate either a circular or a linear DNA or RNA molecule, which is either double-stranded or single-stranded, and which comprise at least one

polynucleotide of interest that is sought to be transferred in a cell host or in a unicellular or multicellular host organism.

The term "primer" denotes a specific oligonucleotide sequence which is complementary to a target nucleotide sequence and used to hybridize to the target nucleotide sequence. A primer serves 5 as an initiation point for nucleotide polymerization catalyzed by either DNA polymerase, RNA polymerase or reverse transcriptase.

The term "probe" denotes a defined nucleic acid segment (or nucleotide analog segment, e.g., polynucleotide as defined hereinbelow) which can be used to identify a specific polynucleotide sequence present in samples, said nucleic acid segment comprising a nucleotide sequence 10 complementary of the specific polynucleotide sequence to be identified.

The terms "trait" and "phenotype" are used interchangeably herein and refer to any visible, detectable or otherwise measurable property of an organism such as symptoms of, or susceptibility to a disease for example.

The term "allele" is used herein to refer to variants of a nucleotide sequence. A biallelic 15 polymorphism has two forms. Diploid organisms may be homozygous or heterozygous for an allelic form.

The term "genotype" as used herein refers the identity of the alleles present in an individual or a sample. In the context of the present invention, a genotype preferably refers to the description 20 of the biallelic marker alleles present in an individual or a sample. The term "genotyping" a sample or an individual for a biallelic marker involves determining the specific allele or the specific nucleotide carried by an individual at a biallelic marker.

The term "mutation" as used herein refers to a difference in DNA sequence between or among different genomes or individuals which has a frequency below 1%.

The term "polymorphism" as used herein refers to the occurrence of two or more alternative 25 genomic sequences or alleles between or among different genomes or individuals. "Polymorphic" refers to the condition in which two or more variants of a specific genomic sequence can be found in a population. A "polymorphic site" is the locus at which the variation occurs. A single nucleotide polymorphism is the replacement of one nucleotide by another nucleotide at the polymorphic site. Deletion of a single nucleotide or insertion of a single nucleotide also gives rise to single nucleotide 30 polymorphisms. In the context of the present invention, "single nucleotide polymorphism" preferably refers to a single nucleotide substitution. Typically, between different individuals, the polymorphic site may be occupied by two different nucleotides.

The term "biallelic polymorphism" and "biallelic marker" are used interchangeably herein to refer to a single nucleotide polymorphism having two alleles at a fairly high frequency in the 35 population. A "biallelic marker allele" refers to the nucleotide variants present at a biallelic marker site.

The location of nucleotides in a polynucleotide with respect to the center of the polynucleotide are described herein in the following manner. When a polynucleotide has an odd number of nucleotides, the nucleotide at an equal distance from the 3' and 5' ends of the polynucleotide is considered to be "at the center" of the polynucleotide, and any nucleotide 5 immediately adjacent to the nucleotide at the center, or the nucleotide at the center itself is considered to be "within 1 nucleotide of the center." With an odd number of nucleotides in a polynucleotide any of the five nucleotides positions in the middle of the polynucleotide would be considered to be within 2 nucleotides of the center, and so on. When a polynucleotide has an even 10 number of nucleotides, there would be a bond and not a nucleotide at the center of the polynucleotide. Thus, either of the two central nucleotides would be considered to be "within 1 nucleotide of the center" and any of the four nucleotides in the middle of the polynucleotide would be considered to be "within 2 nucleotides of the center", and so on.

Biallelic markers can be defined as genome-derived polynucleotides having between 2 and 100, preferably between 20, 30, or 40 and 60, and more preferably about 47 nucleotides in length, 15 which exhibit biallelic polymorphism at one single base position. Each biallelic marker therefore corresponds to two forms of a polynucleotide sequence included in a gene which, when compared with one another, present a nucleotide modification at one position.

The term "upstream" is used herein to refer to a location which is toward the 5' end of the polynucleotide from a specific reference point.

20 The terms "base paired" and "Watson & Crick base paired" are used interchangeably herein to refer to nucleotides which can be hydrogen bonded to one another by virtue of their sequence identities in a manner like that found in double-helical DNA with thymine or uracil residues linked to adenine residues by two hydrogen bonds and cytosine and guanine residues linked by three hydrogen bonds (See Stryer, L., *Biochemistry*, 4th edition, 1995).

25 The terms "complementary" or "complement thereof" are used herein to refer to the sequences of polynucleotides which is capable of forming Watson & Crick base pairing with another specified polynucleotide throughout the entirety of the complementary region. For the purpose of the present invention, a first polynucleotide is deemed to be complementary to a second polynucleotide when each base in the first polynucleotide is paired with its complementary base. Complementary 30 bases are, generally, A and T (or A and U), or C and G. "Complement" is used herein as a synonym from "complementary polynucleotide", "complementary nucleic acid" and "complementary nucleotide sequence". These terms are applied to pairs of polynucleotides based solely upon their sequences and not any particular set of conditions under which the two polynucleotides would actually bind.

Variants and fragments**1- Polynucleotides**

The invention also relates to variants and fragments of the polynucleotides described herein, particularly of an olfactory receptor gene containing one or more biallelic markers according to the 5 invention.

Variants of polynucleotides, as the term is used herein, are polynucleotides that differ from a reference polynucleotide. A variant of a polynucleotide may be a naturally occurring variant such as a naturally occurring allelic variant, or it may be a variant that is not known to occur naturally. Such non-naturally occurring variants of the polynucleotide may be made by mutagenesis techniques, 10 including those applied to polynucleotides, cells or organisms. Generally, differences are limited so that the nucleotide sequences of the reference and the variant are closely similar overall and, in many regions, identical.

Variants of polynucleotides according to the invention include, without being limited to, nucleotide sequences at least 95% identical to a nucleic acid selected from the group consisting of 15 SEQ ID Nos 1-11, or to any polynucleotide fragment of at least 12 consecutive nucleotides from a nucleic acid selected from the group consisting of SEQ ID Nos 1-11, and preferably at least 99% identical, more particularly at least 99.5% identical, and most preferably at least 99.8% identical to a nucleic acid selected from the group consisting of SEQ ID Nos 1-11, or to any polynucleotide fragment of at least 12 consecutive nucleotides from a nucleic acid selected from the group 20 consisting of SEQ ID Nos 1-11.

Changes in the nucleotide of a variant may be silent, which means that they do not alter the amino acids encoded by the polynucleotide. However, nucleotide changes may also result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence. The substitutions, deletions or additions may involve one or more nucleotides. 25 The variants may be altered in coding or non-coding regions or both. Alterations in the coding regions may produce conservative or non-conservative amino acid substitutions, deletions or additions.

In the context of the present invention, particularly preferred embodiments are those in which the polynucleotides encode polypeptides which retain substantially the same biological 30 function or activity as the mature olfactory receptor protein, or those in which the polynucleotides encode polypeptides which maintain or increase a particular biological activity, while reducing a second biological activity.

A polynucleotide fragment is a polynucleotide which sequence is fully comprised within part of a given nucleotide sequence, preferably the nucleotide sequence of an olfactory receptor gene 35 of the invention, and variants thereof. The fragment can be a portion of a coding or non-coding region of the olfactory receptor gene cluster. Preferably, such fragments comprise at least one of the biallelic markers A1 to A13 or the complements thereto or a biallelic marker in linkage

disequilibrium with one or more of the biallelic markers A1 to A13, for which the respective locations in the sequence listing are provided in Table 2.

Such fragments may be "free-standing", i.e. not part of or fused to other polynucleotides, or they may be comprised within a single larger polynucleotide of which they form a part or region.

5 However, several fragments may be comprised within a single larger polynucleotide.

As representative examples of polynucleotide fragments of the invention, there may be mentioned those which have from about 4, 6, 8, 15, 20, 25, 40, 10 to 30, 30 to 55, 50 to 100, 75 to 100 or 100 to 200 nucleotides in length. Preferred are those fragments having about 47 nucleotides in length, such as those comprising at least one of the biallelic markers A1 to A13 of the olfactory 10 receptor gene. Optionally, such fragments may consist of, or consist essentially of a contiguous span of at least 8, 10, 12, 15, 18, 20, 25, 35, 40, 50, 70, 80, 100, 250, 500 or 1000 nucleotides in length. A set of preferred fragments contain at least one of the biallelic markers A1 to A13 of the olfactory receptor gene which are described herein or the complements thereto.

2- Polypeptides

15 The invention also relates to variants, fragments, analogs and derivatives of the polypeptides described herein, including mutated olfactory receptor proteins.

The variant may be 1) one in which one or more of the amino acid residues are substituted with a conserved or non-conserved amino acid residue and such substituted amino acid residue may or may not be one encoded by the genetic code, or 2) one in which one or more of the amino acid 20 residues includes a substituent group, or 3) one in which the mutated olfactory receptor is fused with another compound, such as a compound to increase the half-life of the polypeptide (for example, polyethylene glycol), or 4) one in which the additional amino acids are fused to the mutated olfactory receptor, such as a leader or secretory sequence or a sequence which is employed for purification of the mutated olfactory receptor or a preprotein sequence. Such variants are deemed to 25 be within the scope of those skilled in the art.

In the case of an amino acid substitution in the amino acid sequence of a polypeptide according to the invention, one or several amino acids can be replaced by "equivalent" amino acids. The expression "equivalent" amino acid is used herein to designate any amino acid that may be substituted for one of the amino acids having similar properties, such that one skilled in the art of 30 peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. Generally, the following groups of amino acids represent equivalent changes: (1) Ala, Pro, Gly, Glu, Asp, Gln, Asn, Ser, Thr; (2) Cys, Ser, Tyr, Thr; (3) Val, Ile, Leu, Met, Ala, Phe; (4) Lys, Arg, His; (5) Phe, Tyr, Trp, His.

More particularly, a variant olfactory receptor polypeptide comprises amino acid changes 35 ranging from 1, 2, 3, 4, 5, 10 to 20 substitutions, additions or deletions of one aminoacid, preferably from 1 to 10, more preferably from 1 to 5 and most preferably from 1 to 3 substitutions, additions or deletions of one amino acid. The preferred amino acid changes are those which have little or no

influence on the biological activity or the capacity of the variant olfactory receptor polypeptide to bind to antibodies raised against a native olfactory receptor protein.

A specific, but not restrictive, embodiment of a modified peptide molecule of interest according to the present invention, which consists in a peptide molecule which is resistant to 5 proteolysis, is a peptide in which the -CONH- peptide bond is modified and replaced by a (CH₂NH) reduced bond, a (NHCO) retro inverso bond, a (CH₂-O) methylene-oxy bond, a (CH₂-S) thiomethylene bond, a (CH₂CH₂) carba bond, a (CO-CH₂) cetomethylene bond, a (CHOH-CH₂) hydroxyethylene bond), a (N-N) bound, a E-alcene bond or also a -CH=CH- bond.

The polypeptide according to the invention could have post-translational modifications. For 10 example, it can present the following modifications: acylation, disulfide bond formation, prenylation, carboxymethylation and phosphorylation.

A polypeptide fragment is a polypeptide which sequence is fully comprised within part of a given polypeptide sequence, preferably a polypeptide encoded by an olfactory receptor gene and variants thereof.

15 Such fragments may be “free-standing”, i.e. not part of or fused to other polypeptides, or they may be comprised within a single larger polypeptide of which they form a part or region. However, several fragments may be comprised within a single larger polypeptide.

As representative examples of polypeptide fragments of the invention, there may be mentioned those which have from about 5, 6, 7, 8, 9 or 10 to 15, 10 to 20, 15 to 40, or 30 to 55 20 amino acids long. Preferred polypeptide fragments according to the invention comprise a contiguous span of at least 6 amino acids, preferably at least 8 or amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of one amino acid sequence. Preferred are those fragments containing at least one amino acid mutation in the olfactory receptor protein under consideration.

Identity between nucleic acids or polypeptides

25 The terms “percentage of sequence identity” and “percentage homology” are used interchangeably herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise 30 additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Homology is evaluated using either any of the 35 variety of sequence comparison algorithms and programs known in the art, or by eye inspection. Such algorithms and programs include, but are by no means limited to, TBLASTN, BLASTP, FASTA, TFASTA, and CLUSTALW (Pearson and Lipman, 1988; Altschul et al., 1990; Thompson

et al., 1994; Higgins et al., 1996; Altschul et al., 1990; Altschul et al., 1993). In a particularly preferred embodiment, protein and nucleic acid sequence homologies are evaluated using the Basic Local Alignment Search Tool (“BLAST”) which is well known in the art (see, e.g., Karlin and Altschul, 1990; Altschul et al., 1990, 1993, 1997). In particular, five specific BLAST programs are 5 used to perform the following task:

- (1) BLASTP and BLAST3 compare an amino acid query sequence against a protein sequence database;
- (2) BLASTN compares a nucleotide query sequence against a nucleotide sequence database;
- 10 (3) BLASTX compares the six-frame conceptual translation products of a query nucleotide sequence (both strands) against a protein sequence database;
- (4) TBLASTN compares a query protein sequence against a nucleotide sequence database translated in all six reading frames (both strands); and
- (5) TBLASTX compares the six-frame translations of a nucleotide query sequence against 15 the six-frame translations of a nucleotide sequence database.

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as “high-scoring segment pairs,” between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring 20 matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., 1992; Henikoff and Henikoff, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978). The BLAST programs evaluate the statistical significance of all high-scoring segment pairs identified, and preferably selects those segments which satisfy a user-specified threshold of significance, such as a user- 25 specified percent homology. Preferably, the statistical significance of a high-scoring segment pair is evaluated using the statistical significance formula of Karlin (see, e.g., Karlin and Altschul, 1990). The BLAST programs may be used with the default parameters or with modified parameters provided by the user.

Stringent Hybridization Conditions

30 By way of example and not limitation, procedures using conditions of high stringency are as follows: Prehybridization of filters containing DNA is carried out for 8 h to overnight at 65°C in buffer composed of 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 µg/ml denatured salmon sperm DNA. Filters are hybridized for 48 h at 65°C, the preferred hybridization temperature, in prehybridization mixture containing 100 µg/ml denatured 35 salmon sperm DNA and 5-20 X 10⁶ cpm of ³²P-labeled probe. Alternatively, the hybridization step can be performed at 65°C in the presence of SSC buffer, 1 x SSC corresponding to 0.15M NaCl and 0.05 M Na citrate. Subsequently, filter washes can be done at 37°C for 1 h in a solution containing 2

x SSC, 0.01% PVP, 0.01% Ficoll, and 0.01% BSA, followed by a wash in 0.1 X SSC at 50°C for 45 min. Alternatively, filter washes can be performed in a solution containing 2 x SSC and 0.1% SDS, or 0.5 x SSC and 0.1% SDS, or 0.1 x SSC and 0.1% SDS at 68°C for 15 minute intervals.

Following the wash steps, the hybridized probes are detectable by autoradiography. Other 5 conditions of high stringency which may be used are well known in the art and as cited in Sambrook et al., 1989; and Ausubel et al., 1989. These hybridization conditions are suitable for a nucleic acid molecule of about 20 nucleotides in length. There is no need to say that the hybridization conditions described above are to be adapted according to the length of the desired nucleic acid, following techniques well known to the one skilled in the art. The suitable hybridization conditions may for 10 example be adapted according to the teachings disclosed in the book of Hames and Higgins (1985) or in Sambrook et al.(1989).

HOMOLOGIES OF THE NOVEL OLFACTORY RECEPTOR GENE WITH KNOWN OLFACTORY RECEPTORS

A comparison analysis of various olfactory receptor amino acid sequences, including the 15 novel sequences of the invention, has been performed with the alignment program Pileup and the translation program MAP (Winsconsin Package version 8, GCG). The protein sequences were sorted into different families and subfamilies, taking into account their Amino acid Sequence Identity (ASI). It was observed the Open Reading Frames of the OLF1 to OLF10 genes are genetically clearly distinguished from the already known olfactory receptor sequences. For example, the 20 olfactory receptor OLF2 presents respectively 39.9 %, 43.1 % and 44.2 % of identity with prior art olfactory receptors referred in Genbank as L35475, U58675_1 and Y10530. In addition, the nucleotide sequences of Orf-2 to Orf-10 according to the invention are all grouped together, whereas the nucleotide Orf-1 of the invention forms a new family by itself. These amino acid sequence comparison data clearly indicate that the novel olfactory receptor sequences of the invention share 25 common genetic characteristics (Orf-2 to Orf-10) or have specific characteristics (Orf-1) that are not found in the prior art olfactory receptor sequences.

A. OLF1 TO OLF10 GENE POLYNUCLEOTIDES.

The cluster of ten olfactory receptor genes has been found by the inventors to be located on the human chromosome 11, more precisely within the 11q12-q13 locus of said chromosome as 30 described in Example 1.

1. Genomic sequences of the olfactory receptor gene

The present invention concerns the genomic sequence of an olfactory receptor cluster. The present invention encompasses the olfactory receptor gene, or olfactory receptor genomic sequences consisting of, consisting essentially of, or comprising the sequence of SEQ ID No 1, a sequence

complementary thereto, as well as fragments and variants thereof. These polynucleotides may be purified, isolated, or recombinant.

The invention also encompasses a purified, isolated, or recombinant polynucleotide comprising a nucleotide sequence having at least 70, 75, 80, 85, 90, or 95% nucleotide identity with 5 a nucleotide sequence of SEQ ID No 1 or a complementary sequence thereto or a fragment thereof. The nucleotide differences as regards to the nucleotide sequence of SEQ ID No 1 may be generally randomly distributed throughout the entire nucleic acid. Nevertheless, preferred nucleic acids are those wherein the nucleotide differences as regards to the nucleotide sequence of SEQ ID No 1 are predominantly located outside the coding sequences contained in the exons. These nucleic acids, as 10 well as their fragments and variants, may be used as oligonucleotide primers or probes in order to detect the presence of a copy of the olfactory receptor gene in a test sample, or alternatively in order to amplify a target nucleotide sequence within the olfactory receptor sequences.

Another object of the invention consists of a purified, isolated, or recombinant nucleic acid that hybridizes with the nucleotide sequence of SEQ ID No 1 or a complementary sequence thereto, 15 under stringent hybridization conditions as defined above.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide 20 positions of SEQ ID No 1: 1-113643, 114064-127488, 127855-144460. Additional preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-10000, 25 10001-20000, 20001-30000, 30001-40000, 40001-50000, 50001-60000, 60001-70000, 70001-80000, 80001-90000, 90001-100000, 100001-110000, 110001-120000, 120001-130000, 130001-140000, and 140001-144460. Further preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the 30 complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-5000, 5001-10000, 10001-15000, 15001-20000, 20001-25000, 25001-30000, 30001-35000, 35001-40000, 40001-45000, 45001-50000, 50001-55000, 55001-60000, 60001-65000, 65001-70000, 70001-75000, 75001-80000, 80001-85000, 85001-90000, 90001-95000, 95001-100000, 100001-105000, 105001-110000, 110001-115000, 35 115001-120000, 120001-125000, 125001-130000, 130001-135000, 135001-140000, and 140001-144460.

The olfactory receptor genomic nucleic acid comprises 10 open reading frames, each carried by a single exon and encoding a polypeptide designated OLF1 to OLF10. The open reading frames positions of OLF1 to OLF10 in SEQ ID No 1 are given as features in the sequence listing and are also detailed below in Table A.

5 Two truncated ubiquitin polypeptides Ubi1 and Ubi2, unrelated to olfactory receptor coding sequences, are encoded on the complementary strand of the olfactory receptor gene. The complementary sequence of the Ubi1 ORF is located between the nucleotide in position 114063 and the nucleotide in position 113644 of the nucleotide sequence of SEQ ID No 1. The complementary sequence of the Ubi2 ORF is located between the nucleotide in position 127854 and the nucleotide
10 in position 127489 of the nucleotide sequence of SEQ ID No 1.

Table A

Coding regions			Non-coding regions		
Name	Position in SEQ ID No 1		Name	Position in SEQ ID No 1	
	Beginning	End		Beginning	End
OLF1	2406	2600	NC1	1	2405
OLF2	9711	10658	NC2	2601	9710
OLF3	24851	25369	NC3	10659	24850
OLF4	45714	46661	NC4	25370	45713
OLF5	80198	81115	NC5	46662	80197
OLF6	96291	96902	NC6	81116	96290
OLF7	110758	111564	NC7	96903	110757
OLF8	122525	122887	NC8	111565	122524
OLF9	132454	133389	NC9	122888	132453
OLF10	143398	143577	NC10	133390	143397
			NC11	143578	144460

Thus, the invention embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the 10 open reading frames of the
15 olfactory receptor gene, or a sequence complementary thereto.

The nucleic acid of SEQ ID No 1 also comprises non coding portions flanking each of the ten olfactory receptor open reading frames of the sense DNA strand.

The invention also embodies purified, isolated, or recombinant polynucleotides comprising a nucleotide sequence selected from the group consisting of the non-coding regions contained in the
20 olfactory receptor gene cluster of SEQ ID No 1, or a sequence complementary thereto as well as their fragments or variants. The term "non-coding" sequence refers to any nucleotide sequence which does not encode an amino acid. The non-coding sequences encompass upstream and downstream regions of the olfactory receptor ORFs of the invention, as well as regions located

between two successive olfactory receptor ORFs, as indicated in Table A which lists the 11 non-coding regions named from NC1 to NC11.

The nucleic acids defining the non-coding sequences of the polynucleotide of SEQ ID No 1 described above, as well as their fragments and variants, may be used as oligonucleotide primers or 5 probes in order to detect the presence of a copy of one of the olfactory receptor genes of the invention in a test sample, or alternatively in order to amplify a target nucleotide sequence within the cluster of olfactory receptor encoding sequences according to the invention.

While this section is entitled "Genomic Sequences of the olfactory receptor gene," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the 10 polynucleotides described in this section, flanking the genomic sequences of olfactory receptor on either side or between two or more such genomic sequences.

2. Coding regions of the olfactory receptor gene

The 10 olfactory receptor open reading frames are presented individually as SEQ ID Nos 2-11 in the appended sequence listing.

15 Thus, another object of the invention is a purified, isolated, or recombinant nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 2-11, complementary sequences thereto, as well as allelic variants, and fragments thereof. Moreover, preferred polynucleotides of the invention include purified, isolated, or recombinant olfactory receptor cDNAs consisting of, consisting essentially of, or comprising a sequence selected from the 20 group consisting of SEQ ID Nos 2-11.

The invention also pertains to a purified or isolated nucleic acid comprising a polynucleotide having at least 95% nucleotide identity with a polynucleotide selected from the group consisting of SEQ ID Nos 2-11, advantageously 99 % nucleotide identity, preferably 99.5% nucleotide identity and most preferably 99.8% nucleotide identity with a polynucleotide selected from the group 25 consisting of SEQ ID Nos 2-11, or a sequence complementary thereto or a biologically active fragment thereof.

Another object of the invention relates to purified, isolated or recombinant nucleic acids comprising a polynucleotide that hybridizes, under the stringent hybridization conditions defined herein, with a polynucleotide selected from the group consisting of SEQ ID Nos 2-11, or a sequence 30 complementary thereto or a biologically active fragment thereof.

Particularly preferred nucleic acids of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group 35 consisting of SEQ ID Nos 2-11 or the complements thereof. Additional preferred embodiments of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2-11 or the

complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of said selected sequence : 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 601-650, 651-700, 701-750, 751-800, 801-850, 851-900, 901- the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide positions are consistent with the lengths of the particular olfactory receptor coding region being referred to. Further preferred embodiments of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2, 4, 7, 9 and 11, or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of said selected sequence: 1-25, 26-50, 51-75, 76-100, 101-125, 126-150, 151-175, 176-200, 201-225, 226-250, 251-275, 276-300, 301-325, 326-350, 351-375, 376-400, 401-425, 426-450, 451-475, 476-500, 501-525, 526-550, 551-575, 576-the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide positions are consistent with the lengths of the particular olfactory receptor coding region being referred to.

The present invention also embodies isolated, purified, and recombinant polynucleotides encoding olfactory receptor polypeptides, wherein olfactory receptor polypeptides comprise an amino acid sequence selected from the group consisting of SEQ ID Nos 12-21, a nucleotide sequence complementary thereto, a fragment or a variant thereof. The present invention also embodies isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21. In a preferred embodiment, the present invention embodies isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21 wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the following amino acid positions in said selected sequence: 1-20, 21-40, 41-60, 61-80, 81-100, 101-120, 121-140, 141-160, 161-180, 181-200, 201-220, 221-240, 241-260, 261-280, 281-300, 301-the terminal amino acid of the olfactory receptor proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to. In another preferred embodiment, the present invention embodies isolated, purified, and recombinant polynucleotides which encode polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12, 14, 17, 19 or 21 wherein said contiguous span includes at least 1, 2, 3, 5 or 7 of the following amino acid positions in said selected sequence: 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-

110, 111-120, 121-130, 131-140, 141-150, 151-160, 161-170, 171-180, 181-190, 191-the terminal amino acid of the olfactory receptor proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to.

In further preferred embodiments, the present invention embodies isolated, purified, and 5 recombinant polynucleotides which encode olfactory receptor polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID No 12-21, wherein said contiguous span includes at least one amino acid at the following positions of said selected sequence:

10 i) 1-3, 10, 16, 21, 28, 33, 34, 36, 42-44, 46, 49, 53, 54, 57, 59, 63, and 64 for SEQ ID No 12;

 ii) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 295, 298, 299, 301, 311, and 313-315 for SEQ ID No 13;

15 iii) 2-4, 6, 18, 21, 25, 34, 37, 98, 99, 102, 113, 114, 133, 143, 148, 158-163, 166, 167, 169, and 170 for SEQ ID No 14;

 iv) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 54, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 298, 299, 311, and 313-315 for SEQ ID No 15;

20 v) 3, 18, 20, 25, 34, 47, 49, 67, 97, 100, 107, 108, 112, 113, 126, 135, 142, 146, 147, 157, 159-160, 194, 196, 228, 245, 264, 265, 269, 279, 298, and 302 for SEQ ID No 16;

 vi) 2, 6, 18, 20, 33, 34, 37, 65, 68, 69, 72, 86, 88, 101, 107, 113, 114, 148, 158, 161, 164, 195, and 198 for SEQ ID No 17;

25 vii) 2, 6, 7, 52, 56, 67, 88, 94, 97, 110, 113, 116, 119, 120, 127, 135, 150, 153, 164, 174, 175, 180, 184, 217, 221, 259, 261, and 268 for SEQ ID No 18;

 viii) 17, 18, 20, 28, 33, 35, 49-52, 105, 111, and 112 for SEQ ID No 19;

 ix) 17, 20, 33, 35, 49-53, 56, 111, 112, 132, 138, 141, 147, 154, 157, 160, 163, 164, 194, 197, 204, 211, 214, 218, 219, 252, 265, 286, 295, 301, 303, 305, 306 and 309 for SEQ ID No 20; and

 x) 9, 18, 26-28, 34, 47 and 50 for SEQ ID No 21, to the extent that such amino acid lengths are consistent with the lengths of the particular olfactory receptor protein being referred to.

Additional preferred fragments of the nucleotide sequences of SEQ ID Nos 2-11 are those 35 encoding olfactory receptor polypeptide fragments located outside the transmembrane domains of the corresponding protein as located in boxes in Figure 1.

The above disclosed polynucleotides that contain only coding sequences derived from the olfactory receptor ORFs may be expressed in a desired host cell or a desired host organism, when said polynucleotides are placed under the control of suitable expression signals. Such a polynucleotide, when placed under suitable expression signals, may be inserted in a vector for its 5 expression.

While this section is entitled "Coding regions of the olfactory receptor gene," it should be noted that nucleic acid fragments of any size and sequence may also be comprised by the polynucleotides described in this section, flanking the genomic sequences of olfactory receptor on either side or between two or more such genomic sequences.

10 3. Polynucleotide Constructs

The terms "polynucleotide construct" and "recombinant polynucleotide" are used interchangeably herein to refer to linear or circular, purified or isolated polynucleotides that have been artificially designed and which comprise at least two nucleotide sequences that are not found as contiguous nucleotide sequences in their initial natural environment.

15 DNA Construct That Enables Directing Temporal And Spatial olfactory receptor Gene Expression In Recombinant Cell Hosts And In Transgenic Animals.

In order to study the physiological and phenotypic consequences of a lack of synthesis of the olfactory receptor protein, both at the cell level and at the multi cellular organism level, the invention also encompasses DNA constructs and recombinant vectors enabling a conditional 20 expression of a specific allele of the olfactory receptor genomic sequence or cDNA and also of a copy of this genomic sequence or cDNA harboring substitutions, deletions, or additions of one or more bases as regards to the olfactory receptor nucleotide sequence of SEQ ID Nos 1-11, or a fragment thereof, these base substitutions, deletions or additions being located in the coding regions of the olfactory receptor genomic sequence or within the olfactory receptor open reading frames of 25 SEQ ID Nos 2-11. In a preferred embodiment, the olfactory receptor sequence comprises a biallelic marker of the present invention. In a preferred embodiment, the olfactory receptor sequence comprises a biallelic marker of the present invention, preferably one of the biallelic markers A1 to A13.

The present invention embodies recombinant vectors comprising any one of the 30 polynucleotides described in the present invention. More particularly, the polynucleotide constructs according to the present invention can comprise any of the polynucleotides described in the "Genomic sequences of the olfactory receptor gene" section, the "Coding regions of the olfactory receptor Gene" section, and the "Oligonucleotide probes and primers" section.

DNA Constructs Allowing Homologous Recombination: Replacement Vectors

A first preferred DNA construct will comprise, from 5'-end to 3'-end: (a) a first nucleotide sequence that is comprised in the olfactory receptor genomic sequence; (b) a nucleotide sequence comprising a positive selection marker, such as the marker for neomycin resistance (*neo*); and (c) a 5 second nucleotide sequence that is comprised in the olfactory receptor genomic sequence, and is located on the genome downstream the first olfactory receptor nucleotide sequence (a).

In a preferred embodiment, this DNA construct also comprises a negative selection marker located upstream the nucleotide sequence (a) or downstream the nucleotide sequence (c).

Preferably, the negative selection marker comprises the thymidine kinase (*tk*) gene (Thomas et al., 10 1986), the hygromycin beta gene (Te Riele et al., 1990), the *hprt* gene (Van der Lugt et al., 1991; Reid et al., 1990) or the Diphtheria toxin A fragment (*Dt-A*) gene (Nada et al., 1993; Yagi et al.1990). Preferably, the positive selection marker is located within an olfactory receptor open reading frame sequence so as to interrupt the sequence encoding an olfactory receptor protein. These replacement vectors are described, for example, by Thomas et al.(1986; 1987), Mansour et al.(1988) and Koller 15 et al.(1992).

The first and second nucleotide sequences (a) and (c) may be indifferently located within an olfactory receptor regulatory sequence, an intronic sequence, an exon sequence or a sequence containing both regulatory and/or intronic and/or exon sequences. The size of the nucleotide sequences (a) and (c) ranges from 1 to 50 kb, preferably from 1 to 10 kb, more preferably from 2 to 20 6 kb and most preferably from 2 to 4 kb.

DNA Constructs Allowing Homologous Recombination: Cre-LoxP System.

These new DNA constructs make use of the site specific recombination system of the P1 phage. The P1 phage possesses a recombinase called Cre which interacts specifically with a 34 base pairs *loxP* site. The *loxP* site is composed of two palindromic sequences of 13 bp separated by a 8 25 bp conserved sequence (Hoess et al., 1986). The recombination by the Cre enzyme between two *loxP* sites having an identical orientation leads to the deletion of the DNA fragment.

The Cre-*loxP* system used in combination with a homologous recombination technique has been first described by Gu et al.(1993, 1994). Briefly, a nucleotide sequence of interest to be inserted in a targeted location of the genome harbors at least two *loxP* sites in the same orientation 30 and located at the respective ends of a nucleotide sequence to be excised from the recombinant genome. The excision event requires the presence of the recombinase (Cre) enzyme within the nucleus of the recombinant cell host. The recombinase enzyme may be brought at the desired time either by (a) incubating the recombinant cell hosts in a culture medium containing this enzyme, by injecting the Cre enzyme directly into the desired cell, such as described by Araki et al.(1995), or by 35 lipofection of the enzyme into the cells, such as described by Baubonis et al.(1993); (b) transfecting the cell host with a vector comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter being optionally inducible, said vector being

introduced in the recombinant cell host, such as described by Gu et al.(1993) and Sauer et al.(1988); (c) introducing in the genome of the cell host a polynucleotide comprising the *Cre* coding sequence operably linked to a promoter functional in the recombinant cell host, which promoter is optionally inducible, and said polynucleotide being inserted in the genome of the cell host either by a random 5 insertion event or an homologous recombination event, such as described by Gu et al.(1994).

In a specific embodiment, the vector containing the sequence to be inserted in the olfactory receptor gene by homologous recombination is constructed in such a way that selectable markers are flanked by *loxP* sites of the same orientation, it is possible, by treatment by the *Cre* enzyme, to eliminate the selectable markers while leaving the olfactory receptor sequences of interest that have 10 been inserted by an homologous recombination event. Again, two selectable markers are needed: a positive selection marker to select for the recombination event and a negative selection marker to select for the homologous recombination event. Vectors and methods using the *Cre-loxP* system are described by Zou et al.(1994).

Thus, a second preferred DNA construct of the invention comprises, from 5'-end to 3'-end: 15 (a) a first nucleotide sequence that is comprised in the olfactory receptor genomic sequence; (b) a nucleotide sequence comprising a polynucleotide encoding a positive selection marker, said nucleotide sequence comprising additionally two sequences defining a site recognized by a recombinase, such as a *loxP* site, the two sites being placed in the same orientation; and (c) a second nucleotide sequence that is comprised in the olfactory receptor genomic sequence, and is located on 20 the genome downstream of the first olfactory receptor nucleotide sequence (a).

The sequences defining a site recognized by a recombinase, such as a *loxP* site, are preferably located within the nucleotide sequence (b) at suitable locations bordering the nucleotide sequence for which the conditional excision is sought. In one specific embodiment, two *loxP* sites are located at each side of the positive selection marker sequence, in order to allow its excision at a 25 desired time after the occurrence of the homologous recombination event.

In a preferred embodiment of a method using the third DNA construct described above, the excision of the polynucleotide fragment bordered by the two sites recognized by a recombinase, preferably two *loxP* sites, is performed at a desired time, due to the presence within the genome of the recombinant host cell of a sequence encoding the *Cre* enzyme operably linked to a promoter 30 sequence, preferably an inducible promoter, more preferably a tissue-specific promoter sequence and most preferably a promoter sequence which is both inducible and tissue-specific, such as described by Gu et al.(1994).

The presence of the *Cre* enzyme within the genome of the recombinant cell host may result from the breeding of two transgenic animals, the first transgenic animal bearing the olfactory 35 receptor-derived sequence of interest containing the *loxP* sites as described above and the second transgenic animal bearing the *Cre* coding sequence operably linked to a suitable promoter sequence, such as described by Gu et al.(1994).

Spatio-temporal control of the Cre enzyme expression may also be achieved with an adenovirus based vector that contains the Cre gene thus allowing infection of cells, or *in vivo* infection of organs, for delivery of the Cre enzyme, such as described by Anton and Graham (1995) and Kanegae et al.(1995).

5 The DNA constructs described above may be used to introduce a desired nucleotide sequence of the invention, preferably an olfactory receptor genomic sequence or an olfactory receptor coding region sequences, and most preferably an altered copy of an olfactory receptor genomic or coding region sequences, within a predetermined location of the targeted genome, leading either to the generation of an altered copy of a targeted gene (knock-out homologous 10 recombination) or to the replacement of a copy of the targeted gene by another copy sufficiently homologous to allow an homologous recombination event to occur (knock-in homologous recombination). In a specific embodiment, the DNA constructs described above may be used to introduce an olfactory receptor genomic sequence or an olfactory receptor coding region sequence comprising at least one biallelic marker of the present invention, preferably at least one biallelic 15 marker selected from the group consisting of A1 to A13.

Nuclear Antisense DNA Constructs

Other compositions containing a vector of the invention comprising an oligonucleotide fragment of the nucleic sequence SEQ ID Nos 2-11, preferably a fragment including the start codon of the olfactory receptor gene, as an antisense tool that inhibits the expression of the corresponding 20 olfactory receptor gene. Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995) or those described in PCT Application No WO 95/24223.

Preferred antisense polynucleotides according to the present invention are complementary to a sequence of the mRNAs of olfactory receptor that contains the translation initiation codon ATG. 25 Preferably, the antisense polynucleotides of the invention have a 3' polyadenylation signal that has been replaced with a self-cleaving ribozyme sequence, such that RNA polymerase II transcripts are produced without poly(A) at their 3' ends, these antisense polynucleotides being incapable of export from the nucleus, such as described by Liu et al.(1994). In a preferred embodiment, these olfactory receptor antisense polynucleotides also comprise, within the ribozyme 30 cassette, a histone stem-loop structure to stabilize cleaved transcripts against 3'-5' exonuclease degradation, such as the structure described by Eckner et al.(1991).

4. Oligonucleotide probes and primers

Polynucleotides derived from the olfactory receptor gene are useful in order to detect the presence of at least a copy of a nucleotide sequence of SEQ ID Nos 1-11, or a fragment, 35 complement, or variant thereof in a test sample, preferably a human olfactory epithelium tissue or isolated human olfactory epithelium cells.

Particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide 5 positions of SEQ ID No 1: 1-113643, 114064-127488, 127855-144460. Additional preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-10000, 10001-20000, 20001-30000, 30001-40000, 40001-50000, 50001-60000, 60001-70000, 70001- 10 80000, 80001-90000, 90001-100000, 100001-110000, 110001-120000, 120001-130000, 130001- 140000, and 140001-144460. Further preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of 12, 15, 18, 20, 25, 30, 35, 15 40, 50, 60, 70, 80, 90, 100, 150, 200, 500, or 1000 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of SEQ ID No 1: 1-5000, 5001-10000, 10001-15000, 15001-20000, 20001-25000, 25001- 30000, 30001-35000, 35001-40000, 40001-45000, 45001-50000, 50001-55000, 55001-60000, 60001-65000, 65001-70000, 70001-75000, 75001-80000, 80001-85000, 85001-90000, 90001- 15 95000, 95001-100000, 100001-105000, 105001-110000, 110001-115000, 115001-120000, 120001- 20 125000, 125001-130000, 130001-135000, 135001-140000, and 140001-144460.

Other particularly preferred probes and primers of the invention include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 25, 30, 35, 40, 45 or 50 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2-11 or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the 25 following nucleotide positions of said selected sequence : 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 601-650, 651-700, 701-750, 751- 800, 801-850, 851-900, 901- the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide positions are consistent with the lengths of the particular olfactory receptor coding region being referred to. Further preferred probes and primers of the invention 30 include isolated, purified, or recombinant polynucleotides comprising a contiguous span of at least 12, 15, 18, 20, 22 or 25 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2, 4, 7, 9 and 11, or the complements thereof, wherein said contiguous span comprises at least 1, 2, 3, 5, or 10 of the following nucleotide positions of said selected sequence: 1-25, 26-50, 51-75, 76- 100, 101-125, 126-150, 151-175, 176-200, 201-225, 226-250, 251-275, 276-300, 301-325, 326-350, 35 351-375, 376-400, 401-425, 426-450, 451-475, 476-500, 501--525, 526-550, 551-575, 576-the terminal nucleotide of the olfactory receptor coding regions, to the extent that such nucleotide

positions are consistent with the lengths of the particular olfactory receptor coding region being referred to.

Thus, the invention also relates to nucleic acid probes characterized in that they hybridize specifically, under the stringent hybridization conditions defined above, with a nucleic acid selected 5 from the group consisting of SEQ ID Nos 1-11, a variant thereof and a sequence complementary thereto.

In one embodiment the invention encompasses isolated, purified, and recombinant polynucleotides consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 and the complement thereof, wherein said span includes an olfactory receptor-related 10 biallelic marker in said sequence; optionally, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof; optionally, wherein said contiguous span is 18 to 47 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide; optionally, wherein said polynucleotide consists of said contiguous span and said contiguous span is 25 nucleotides in length and said biallelic marker is 15 at the center of said polynucleotide; optionally, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide; and optionally, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide. In a preferred embodiment, said probes comprises, consists of, or consists 20 essentially of a sequence selected from the following sequences: P1 to P13 and the complementary sequences thereto, for which the respective locations in the sequence listing are provided in Table 3.

In another embodiment the invention encompasses isolated, purified and recombinant polynucleotides comprising, consisting of, or consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1, or the complements thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said polynucleotide is 25 located within 20 nucleotides upstream of an olfactory receptor-related biallelic marker in said sequence; optionally, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof; optionally, wherein the 3' end of said polynucleotide is located 1 nucleotide upstream of said olfactory receptor-related biallelic marker in said sequence; and optionally, wherein said polynucleotide consists essentially of a sequence 30 selected from the following sequences: D1 to D13 and E1 to E13, for which the respective locations in the sequence listing are provided in Table 4.

In a further embodiment, the invention encompasses isolated, purified, or recombinant polynucleotides comprising, consisting of, or consisting essentially of a sequence selected from the following sequences: B1 to B11 and C1 to C11, for which the respective locations in the sequence 35 listing are provided in Table 1.

In an additional embodiment, the invention encompasses polynucleotides for use in hybridization assays, sequencing assays, and enzyme-based mismatch detection assays for

determining the identity of the nucleotide at an olfactory receptor-related biallelic marker in SEQ ID No 1, or the complements thereof, as well as polynucleotides for use in amplifying segments of nucleotides comprising an olfactory receptor-related biallelic marker in SEQ ID No 1, or the complements thereof; optionally, wherein said olfactory receptor-related biallelic marker is selected 5 from the group consisting of A1 to A13, and the complements thereof.

A probe or a primer according to the invention has between 8 and 1000 nucleotides in length, or is specified to be at least 12, 15, 18, 20, 25, 35, 40, 50, 60, 70, 80, 100, 250, 500 or 1000 nucleotides in length. More particularly, the length of these probes and primers can range from 8, 10, 15, 20, or 30 to 100 nucleotides, preferably from 10 to 50, more preferably from 15 to 30 10 nucleotides. Shorter probes and primers tend to lack specificity for a target nucleic acid sequence and generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. Longer probes and primers are expensive to produce and can sometimes self-hybridize to form hairpin structures. The appropriate length for primers and probes under a particular set of assay conditions may be empirically determined by one of skill in the art. A preferred probe or 15 primer consists of a nucleic acid comprising a polynucleotide selected from the group of the nucleotide sequences of P1 to P13 and the complementary sequence thereto, B1 to B11, C1 to C11, D1 to D13, and E1 to E13.

Primers and other oligonucleotides according to the invention are synthesized to be "substantially" complementary to a strand of the olfactory receptor gene of the invention to be 20 amplified. The primer sequence does not need to reflect the exact sequence of the DNA template. Minor mismatches can be accommodated by reducing the stringency of the hybridization conditions. Among the various methods available to design useful primers, the OSP computer software can be used by the skilled person (see Hillier & Green, 1991). All primers contained a common upstream 25 oligonucleotide tail enabling the easy systematic sequencing of the resulting amplification fragments.

The formation of stable hybrids depends on the melting temperature (Tm) of the DNA. The Tm depends on the length of the primer or probe, the ionic strength of the solution and the G+C content. The higher the G+C content of the primer or probe, the higher is the melting temperature because G:C pairs are held by three H bonds whereas A:T pairs have only two. The GC content in 30 the probes of the invention usually ranges between 10 and 75 %, preferably between 35 and 60 %, and more preferably between 40 and 55 %.

The primers and probes can be prepared by any suitable method, including, for example, cloning and restriction of appropriate sequences and direct chemical synthesis by a method such as the phosphodiester method of Narang et al.(1979), the phosphodiester method of Brown et al.(1979), 35 the diethylphosphoramide method of Beaucage et al.(1981) and the solid support method described in EP 0 707 592.

Detection probes are generally nucleic acid sequences or uncharged nucleic acid analogs such as, for example peptide nucleic acids which are disclosed in International Patent Application WO 92/20702, morpholino analogs which are described in U.S. Patents Numbered 5,185,444; 5,034,506 and 5,142,047. The probe may have to be rendered "non-extendable" in that additional 5 dNTPs cannot be added to the probe. In and of themselves analogs usually are non-extendable and nucleic acid probes can be rendered non-extendable by modifying the 3' end of the probe such that the hydroxyl group is no longer capable of participating in elongation. For example, the 3' end of the probe can be functionalized with the capture or detection label to thereby consume or otherwise block the hydroxyl group. Alternatively, the 3' hydroxyl group simply can be cleaved, replaced or 10 modified, U.S. Patent Application Serial No. 07/049,061 filed April 19, 1993 describes modifications, which can be used to render a probe non-extendable.

Any of the polynucleotides of the present invention can be labeled, if desired, by incorporating any label known in the art to be detectable by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive 15 substances (including, ³²P, ³⁵S, ³H, ¹²⁵I), fluorescent dyes (including, 5-bromodesoxyuridin, fluorescein, acetylaminofluorene, digoxigenin) or biotin. Preferably, polynucleotides are labeled at their 3' and 5' ends. Examples of non-radioactive labeling of nucleic acid fragments are described in the French patent No. FR-7810975 or by Urdea et al (1988) or Sanchez-Pescador et al (1988). In addition, the probes according to the present invention may have structural characteristics such that 20 they allow the signal amplification, such structural characteristics being, for example, branched DNA probes as those described by Urdea et al. in 1991 or in the European patent No. EP 0 225 807 (Chiron).

A label can also be used to capture the primer, so as to facilitate the immobilization of either the primer or a primer extension product, such as amplified DNA, on a solid support. A capture 25 label is attached to the primers or probes and can be a specific binding member which forms a binding pair with the solid's phase reagent's specific binding member (e.g. biotin and streptavidin). Therefore depending upon the type of label carried by a polynucleotide or a probe, it may be employed to capture or to detect the target DNA. Further, it will be understood that the polynucleotides, primers or probes provided herein, may, themselves, serve as the capture label. For 30 example, in the case where a solid phase reagent's binding member is a nucleic acid sequence, it may be selected such that it binds a complementary portion of a primer or probe to thereby immobilize the primer or probe to the solid phase. In cases where a polynucleotide probe itself serves as the binding member, those skilled in the art will recognize that the probe will contain a sequence or "tail" that is not complementary to the target. In the case where a polynucleotide primer 35 itself serves as the capture label, at least a portion of the primer will be free to hybridize with a nucleic acid on a solid phase. DNA Labeling techniques are well known to the skilled technician.

The probes of the present invention are useful for a number of purposes. They can be notably used in Southern hybridization to genomic DNA or Northern hybridization to mRNA. The probes can also be used to detect PCR amplification products. They may also be used to detect mismatches in the OLF1 to OLF10 genes or mRNA using other techniques. Generally, the probes 5 are complementary to the OLF1 to OLF10 gene coding sequences, although probes complementary to non-coding sequences are also contemplated. The probes of the present invention can also be useful for genotyping the biallelic markers of the cluster of olfactory receptor genes of the present invention.

Any of the polynucleotides, primers and probes of the present invention can be conveniently 10 immobilized on a solid support. Solid supports are known to those skilled in the art and include the walls of wells of a reaction tray, test tubes, polystyrene beads, magnetic beads, nitrocellulose strips, membranes, microparticles such as latex particles, sheep (or other animal) red blood cells, duracytes and others. The solid support is not critical and can be selected by one skilled in the art. Thus, latex particles, microparticles, magnetic or non-magnetic beads, membranes, plastic tubes, walls of 15 microtiter wells, glass or silicon chips, sheep (or other suitable animal's) red blood cells and duracytes are all suitable examples. Suitable methods for immobilizing nucleic acids on solid phases include ionic, hydrophobic, covalent interactions and the like. A solid support, as used herein, refers to any material which is insoluble, or can be made insoluble by a subsequent reaction. The solid support can be chosen for its intrinsic ability to attract and immobilize the capture reagent.

20 Alternatively, the solid phase can retain an additional receptor which has the ability to attract and immobilize the capture reagent. The additional receptor can include a charged substance that is oppositely charged with respect to the capture reagent itself or to a charged substance conjugated to the capture reagent. As yet another alternative, the receptor molecule can be any specific binding member which is immobilized upon (attached to) the solid support and which has the ability to 25 immobilize the capture reagent through a specific binding reaction. The receptor molecule enables the indirect binding of the capture reagent to a solid support material before the performance of the assay or during the performance of the assay. The solid phase thus can be a plastic, derivatized plastic, magnetic or non-magnetic metal, glass or silicon surface of a test tube, microtiter well, sheet, bead, microparticle, chip, sheep (or other suitable animal's) red blood cells, duracytes® and other 30 configurations known to those of ordinary skill in the art. The polynucleotides of the invention can be attached to or immobilized on a solid support individually or in groups of at least 2, 5, 8, 10, 12, 15, 20, or 25 distinct polynucleotides of the invention to a single solid support. In addition, polynucleotides other than those of the invention may be attached to the same solid support as one or 35 more polynucleotides of the invention.

Consequently, the invention also comprises a method for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1-11, a

fragment or a variant thereof and a complementary sequence thereto in a sample, said method comprising the following steps of:

a) bringing into contact a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence selected from the group consisting of the nucleotide sequences

5 of SEQ ID Nos 1-11, a fragment or a variant thereof and a complementary sequence thereto and the sample to be assayed; and

b) detecting the hybrid complex formed between the probe and a nucleic acid in the sample.

The invention further concerns a kit for detecting the presence of a nucleic acid comprising a nucleotide sequence selected from a group consisting of SEQ ID Nos 1-11, a fragment or a variant

10 thereof and a complementary sequence thereto in a sample, said kit comprising:

a) a nucleic acid probe or a plurality of nucleic acid probes which can hybridize with a nucleotide sequence selected from the group consisting of the nucleotide sequences of SEQ ID Nos 1-11, a fragment or a variant thereof and a complementary sequence thereto; and

b) optionally, the reagents necessary for performing the hybridization reaction.

15 In a first preferred embodiment of this detection method and kit, said nucleic acid probe or the plurality of nucleic acid probes are labeled with a detectable molecule. In a second preferred embodiment of said method and kit, said nucleic acid probe or the plurality of nucleic acid probes has been immobilized on a substrate. In a third preferred embodiment, the nucleic acid probe or the plurality of nucleic acid probes comprise either a sequence which is selected from the group

20 consisting of the nucleotide sequences of P1 to P13 and the complementary sequence thereto, B1 to B11, C1 to C11, D1 to D13, E1 to E13 or a biallelic marker selected from the group consisting of A1 to A13 and the complements thereto.

Oligonucleotide arrays

A substrate comprising a plurality of oligonucleotide primers or probes of the invention may 25 be used either for detecting or amplifying targeted sequences in the olfactory receptor gene and may also be used for detecting mutations in the coding or in the non-coding sequences of the olfactory receptor gene.

Any polynucleotide provided herein may be attached in overlapping areas or at random 30 locations on the solid support. Alternatively the polynucleotides of the invention may be attached in an ordered array wherein each polynucleotide is attached to a distinct region of the solid support which does not overlap with the attachment site of any other polynucleotide. Preferably, such an ordered array of polynucleotides is designed to be "addressable" where the distinct locations are recorded and can be accessed as part of an assay procedure. Addressable polynucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a 35 substrate in different known locations. The knowledge of the precise location of each polynucleotides location makes these "addressable" arrays particularly useful in hybridization assays. Any addressable array technology known in the art can be employed with the

polynucleotides of the invention. One particular embodiment of these polynucleotide arrays is known as the Genechips™, and has been generally described in US Patent 5,143,854; PCT publications WO 90/15070 and 92/10092. These arrays may generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination 5 of photolithographic methods and solid phase oligonucleotide synthesis (Fodor et al., 1991). The immobilization of arrays of oligonucleotides on solid supports has been rendered possible by the development of a technology generally identified as "Very Large Scale Immobilized Polymer Synthesis" (VLSIPS™) in which, typically, probes are immobilized in a high density array on a solid surface of a chip. Examples of VLSIPS™ technologies are provided in US Patents 5,143,854; 10 and 5,412,087 and in PCT Publications WO 90/15070, WO 92/10092 and WO 95/11995, which describe methods for forming oligonucleotide arrays through techniques such as light-directed synthesis techniques. In designing strategies aimed at providing arrays of nucleotides immobilized on solid supports, further presentation strategies were developed to order and display the oligonucleotide arrays on the chips in an attempt to maximize hybridization patterns and sequence 15 information. Examples of such presentation strategies are disclosed in PCT Publications WO 94/12305, WO 94/11530, WO 97/29212 and WO 97/31256.

In another embodiment of the oligonucleotide arrays of the invention, an oligonucleotide probe matrix may advantageously be used to detect mutations occurring in the olfactory receptor gene. For this particular purpose, probes are specifically designed to have a nucleotide sequence 20 allowing their hybridization to the genes that carry known mutations (either by deletion, insertion or substitution of one or several nucleotides). By known mutations, it is meant, mutations on the olfactory receptor gene that have been identified according to, for example, the technique used by Huang et al.(1996) or Samson et al.(1996).

Another technique that is used to detect mutations in the olfactory receptor gene is the use of 25 a high-density DNA array. Each oligonucleotide probe constituting a unit element of the high density DNA array is designed to match a specific subsequence of the olfactory receptor genomic DNA or cDNA. Thus, an array consisting of oligonucleotides complementary to subsequences of the target gene sequence is used to determine the identity of the target sequence with the wild gene sequence, measure its amount, and detect differences between the target sequence and the reference 30 wild gene sequence of the olfactory receptor gene. In one such design, termed 4L tiled array, is implemented a set of four probes (A, C, G, T), preferably 15-nucleotide oligomers. In each set of four probes, the perfect complement will hybridize more strongly than mismatched probes. Consequently, a nucleic acid target of length L is scanned for mutations with a tiled array containing 4L probes, the whole probe set containing all the possible mutations in the known wild reference 35 sequence. The hybridization signals of the 15-mer probe set tiled array are perturbed by a single base change in the target sequence. As a consequence, there is a characteristic loss of signal or a

“footprint” for the probes flanking a mutation position. This technique was described by Chee et al. in 1996.

Consequently, the invention concerns an array of nucleic acid molecules comprising at least one polynucleotide described above as probes and primers. Preferably, the invention concerns an 5 array of nucleic acid comprising at least two polynucleotides described above as probes and primers.

A further object of the invention consists of an array of nucleic acid sequences comprising either at least one of the sequences selected from the group consisting of P1 to P13, B1 to B11, C1 to C11, D1 to D13, E1 to E13, the sequences complementary thereto, a fragment thereof of at least 8, 10, 12, 15, 18, 20, 25, 30, or 40 consecutive nucleotides thereof, and at least one sequence 10 comprising a biallelic marker selected from the group consisting of A1 to A13 and the complements thereto.

The invention also pertains to an array of nucleic acid sequences comprising either at least two of the sequences selected from the group consisting of P1 to P13, B1 to B11, C1 to C11, D1 to D13, E1 to E13, the sequences complementary thereto, a fragment thereof of at least 8 consecutive 15 nucleotides thereof, and at least two sequences comprising a biallelic marker selected from the group consisting of A1 to A13 and the complements thereof.

B. OLF1 TO OFL10 PROTEINS AND POLYPEPTIDE FRAGMENTS

The proteins encoded by the Open Reading Frames of the OLF1 to OLF10 genes are listed individually in the sequence listing as SEQ ID Nos 12-21.

20 The term “olfactory receptor polypeptides” is used herein to embrace all of the proteins and polypeptides of the present invention. Also forming part of the invention are polypeptides encoded by the polynucleotides of the invention, as well as fusion polypeptides comprising such polypeptides. The invention embodies olfactory receptor proteins from humans, including isolated or purified olfactory receptor proteins consisting of, consisting essentially of, or comprising the 25 sequences of SEQ ID Nos 12-21 or naturally-occurring variants or fragments thereof.

The present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 12-21. In a preferred embodiment, the present invention embodies isolated, purified, and recombinant polypeptides 30 comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of SEQ ID Nos 12-21 wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the following amino acid positions in SEQ ID Nos 12-21: 1-20, 21-40, 41-60, 61-80, 81-100, 101-120, 121-140, 141-160, 161-180, 181-200, 201-220, 221-240, 241-260, 261-280, 281-300, 301-the terminal amino acid of the olfactory receptor 35 proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to. In another preferred embodiment, the present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6

amino acids, preferably at least 8 to 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12, 14, 17, 19 and 21 wherein said contiguous span includes at least 1, 2, 3, 5 or 10 of the following amino acid positions of said selected sequence: 1-10, 11-20, 21-30, 31-40, 41-50, 51-60, 61-70, 71-80, 81-90, 91-100, 101-110, 111-120, 121-130, 131-140, 141-150, 151-160, 161-170, 171-180, 181-190, 191-195, the terminal amino acid of the olfactory receptor proteins, to the extent that such amino acid positions are consistent with the lengths of the particular olfactory receptor protein being referred to. In further preferred embodiments, the present invention embodies isolated, purified, and recombinant polypeptides comprising a contiguous span of at least 6 amino acids, preferably at least 8 or 10 amino acids, more preferably at least 12, 15, 20, 25, 30, 40, 50, or 100 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21, wherein said contiguous span includes at least one amino acid at the following positions of said selected sequence

1) 1-3, 10, 16, 21, 28, 33, 34, 36, 42-44, 46, 49, 53, 54, 57, 59, 63, and 64 for SEQ ID No 12;

2) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 295, 298, 299, 301, 311, and 313-315 for SEQ ID No 13;

3) 2-4, 6, 18, 21, 25, 34, 37, 98, 99, 102, 113, 114, 133, 143, 148, 158-163, 166, 167, 169, and 170 for SEQ ID No 14;

4) 2, 4, 6, 8, 18, 25, 34, 37, 44, 52, 54, 56, 80, 83, 89, 98, 101, 102, 113, 114, 117, 120, 139, 148, 158, 186, 195, 212, 219, 247, 266, 270, 280, 298, 299, 311, and 313-315 for SEQ ID No 15;

5) 3, 18, 20, 25, 34, 47, 49, 67, 97, 100, 107, 108, 112, 113, 126, 135, 142, 146, 147, 157, 159-160, 194, 196, 228, 245, 264, 265, 269, 279, 298, and 302 for SEQ ID No 16;

6) 2, 6, 18, 20, 33, 34, 37, 65, 68, 69, 72, 86, 88, 101, 107, 113, 114, 148, 158, 161, 164, 195, and 198 for SEQ ID No 17;

7) 2, 6, 7, 52, 56, 67, 88, 94, 97, 110, 113, 116, 119, 120, 127, 135, 150, 153, 164, 174, 175, 180, 184, 217, 221, 259, 261, and 268 for SEQ ID No 18;

8) 17, 18, 20, 28, 33, 35, 49-52, 105, 111, and 112 for SEQ ID No 19;

9) 17, 20, 33, 35, 49-53, 56, 111, 112, 132, 138, 141, 147, 154, 157, 160, 163, 164, 194, 197, 204, 211, 214, 218, 219, 252, 265, 286, 295, 301, 303, 305, 306 and 309 for SEQ ID No 20; and

10) 9, 18, 26-28, 34, 47 and 50 for SEQ ID No 21, to the extent that such amino acid lengths are consistent with the lengths of the particular olfactory receptor protein being referred to.

Other preferred OLF1 to OLF10 polypeptide fragments are those located outside the transmembrane domains, most preferably peptide fragments naturally exposed on the cell membrane, particularly those that are available for binding to ligand molecules, either odorant substances or molecules or antibodies directed to the olfactory receptor polypeptides of the 5 invention. Such transmembrane domains TM1 to TM7 are boxed in Figure 1. In other preferred embodiments the contiguous stretch of amino acids comprises the site of a mutation or functional mutation, including a deletion, addition, swap or truncation of the amino acids in the olfactory receptor protein sequence.

The invention also encompasses a purified, isolated, or recombinant polypeptides 10 comprising an amino acid sequence having at least 70, 75, 80, 85, 90, 95, 98 or 99% amino acid identity with the amino acid sequence of SEQ ID Nos 12-21 or a fragment thereof.

The invention also encompasses an olfactory receptor polypeptide or a fragment or a variant thereof in which at least one peptide bound has been modified as defined in the "Definitions" section.

15 A further object of the invention concerns a purified or isolated polypeptide which is encoded by a nucleic acid comprising a nucleotide sequence selected from the group consisting of SEQ ID Nos 1-11 or fragment or variants thereof.

Such mutated olfactory receptor proteins may be the target of diagnostic tools, such as 20 specific monoclonal or polyclonal antibodies, useful for the detecting the mutated olfactory receptor proteins in a sample.

Olfactory receptor proteins are preferably isolated from human or mammalian tissue samples or expressed from human or mammalian genes.

The olfactory receptor polypeptides of the invention is extracted from cells or tissues of 25 humans or non-human animals. Methods for purifying proteins are known in the art, and include the use of detergents or chaotropic agents to disrupt particles followed by differential extraction and separation of the polypeptides by ion exchange chromatography, affinity chromatography, sedimentation according to density, and gel electrophoresis.

In addition, shorter protein fragments may also be prepared by the conventional methods of 30 chemical synthesis, either in a homogenous solution or in solid phase. As an illustrative embodiment of such chemical polypeptide synthesis techniques, it may be cited the homogenous solution technique described by Houbenweyl in 1974. For solid phase synthesis the technique described by Merrifield (1965) may be used in particular.

Alternatively, the proteins of the invention can be made using routine expression methods 35 known in the art as described below and in the section "Expression of a OLF1 to OLF10 coding polynucleotide". Briefly, the polynucleotide encoding the desired polypeptide, is ligated into an expression vector suitable for any convenient host. Both eukaryotic and prokaryotic host systems is used in forming recombinant polypeptides. The polypeptide is then isolated from lysed cells or from

the culture medium and purified to the extent needed for its intended use. Purification is by any technique known in the art, for example, differential extraction, salt fractionation, chromatography, centrifugation, and the like. See, for example, Methods in Enzymology for a variety of methods for purifying proteins.

5 Any olfactory receptor cDNA, including SEQ ID Nos 12-21, may be used to express olfactory receptor proteins and polypeptides. The nucleic acid encoding the olfactory receptor protein or polypeptide to be expressed is operably linked to a promoter in an expression vector using conventional cloning technology. The olfactory receptor insert in the expression vector may comprise the full coding sequence for the olfactory receptor protein or a portion thereof. For example, the olfactory receptor 10 derived insert may encode a polypeptide comprising at least 10 consecutive amino acids of the olfactory receptor protein of SEQ ID Nos 12-21, including any of the polypeptide fragment defined in this section.

The expression vector is any of the mammalian, yeast, insect or bacterial expression systems known in the art. Commercially available vectors and expression systems are available from a variety 15 of suppliers including Genetics Institute (Cambridge, MA), Stratagene (La Jolla, California), Promega (Madison, Wisconsin), and Invitrogen (San Diego, California). If desired, to enhance expression and facilitate proper protein folding, the codon context and codon pairing of the sequence is optimized for the particular expression organism in which the expression vector is introduced, as explained by Hatfield, et al., U.S. Patent No. 5,082,767.

20 In one embodiment, the entire coding sequence of the olfactory receptor cDNA through the poly A signal of the cDNA are operably linked to a promoter in the expression vector. Alternatively, if the nucleic acid encoding a portion of the olfactory receptor protein lacks a methionine to serve as the initiation site, an initiating methionine can be introduced next to the first codon of the nucleic acid using conventional techniques. Similarly, if the insert from the olfactory receptor cDNA lacks a poly A 25 signal, this sequence can be added to the construct by, for example, splicing out the Poly A signal from pSG5 (Stratagene) using BglII and SalI restriction endonuclease enzymes and incorporating it into the mammalian expression vector pXT1 (Stratagene).

The ligated product is transfected into mouse NIH 3T3 cells using Lipofectin (Life Technologies, Inc., Grand Island, New York) under conditions outlined in the product specification. 30 Positive transfectants are selected after growing the transfected cells in 600ug/ml G418 (Sigma, St. Louis, Missouri).

The above procedures may also be used to express a mutant olfactory receptor protein responsible for a detectable phenotype or a portion thereof.

Purification of the recombinant protein or peptide according to the present invention may be 35 realized by passage onto a Nickel or Copper affinity chromatography column. The Nickel chromatography column may contain the Ni-NTA resin (Porath et al., 1975). The polypeptides or peptides thus obtained may be purified, for example by high performance liquid chromatography,

such as reverse phase and/or cationic exchange HPLC, as described by Rougeot et al. (1994). The reason to prefer this kind of peptide or protein purification is the lack of side products found in the elution samples which renders the resultant purified protein or peptide more suitable for a therapeutic use.

5 The expressed protein may also be purified using other conventional purification techniques such as ammonium sulfate precipitation or chromatographic separation based on size or charge. The protein encoded by the nucleic acid insert may also be purified using standard immunochromatography techniques. In such procedures, polyclonal or monoclonal antibodies capable of specifically binding to the expressed olfactory receptor protein of SEQ ID Nos 12-21, or a fragment or a variant thereof, have
10 been previously immobilized onto a chromatography matrix. Such antibodies are described in the section "Antibodies that bind olfactory receptor polypeptides" below. Then, a solution containing the expressed olfactory receptor protein or portion thereof, such as a cell extract, is applied to the chromatography column in conditions allowing the expressed protein to bind to the antibodies in the immunochromatography column. Thereafter, the column is washed to remove non-specifically bound
15 proteins. The specifically bound expressed protein is then released from the column and recovered using standard techniques.

If antibody production is not possible, the nucleic acids encoding the olfactory receptor protein or a portion thereof is incorporated into expression vectors designed for use in purification schemes employing chimeric polypeptides. In such strategies the nucleic acid encoding the olfactory receptor
20 protein or a portion thereof is inserted in frame with the gene encoding the other half of the chimera. The other half of the chimera is β-globin or a nickel binding polypeptide encoding sequence. A chromatography matrix having antibody to β-globin or nickel attached thereto is then used to purify the chimeric protein. Protease cleavage sites are engineered between the β-globin gene or the nickel binding polypeptide and the olfactory receptor protein or portion thereof. Thus, the two polypeptides of the
25 chimera are separated from one another by protease digestion.

One useful expression vector for generating β-globin chimeric proteins is pSG5 (Stratagene), which encodes rabbit β-globin. Intron II of the rabbit β-globin gene facilitates splicing of the expressed transcript, and the polyadenylation signal incorporated into the construct increases the level of expression. These techniques are well known to those skilled in the art of molecular biology. Standard
30 methods are published in methods texts such as Davis et al., (1986) and many of the methods are available from Stratagene, Life Technologies, Inc., or Promega. Polypeptide may additionally be produced from the construct using in vitro translation systems such as the In vitro Express™ Translation Kit (Stratagene).

To confirm expression of the olfactory receptor protein or a portion thereof, the proteins
35 expressed from host cells containing an expression vector containing an insert encoding the olfactory receptor protein or a portion thereof can be compared to the proteins expressed in host cells containing the expression vector without an insert. The presence of a band in samples from cells containing the

expression vector with an insert which is absent in samples from cells containing the expression vector without an insert indicates that the olfactory receptor protein or a portion thereof is being expressed.

Generally, the band will have the mobility expected for the olfactory receptor protein or portion thereof. However, the band may have a mobility different than that expected as a result of modifications such as 5 glycosylation, ubiquitination, or enzymatic cleavage.

Other suitable techniques for producing and purifying the olfactory receptor proteins of the invention or their fragments or variants are also described under the heading "Methods for screening substances or molecules interacting with an olfactory receptor protein".

Thus, the present invention also concerns a method for the producing a polypeptide of the 10 invention, and especially a polypeptide selected from the group of SEQ ID Nos 12-21 or a fragment or a variant thereof, wherein said methods comprises the steps of :

a) culturing, in an appropriate culture medium, a cell host previously transformed or transfected with the recombinant vector comprising a nucleic acid encoding an olfactory receptor polypeptide of the invention, or a fragment or a variant thereof;

15 b) harvesting the culture medium thus conditioned or lyze the cell host, for example by sonication or by an osmotic shock;

c) separating or purifying, from the said culture medium, or from the pellet of the resultant host cell lysate the thus produced polypeptide of interest.

d) optionally characterizing the produced polypeptide of interest.

20 In a specific embodiment of the above method, step a) is preceded by a step wherein the nucleic acid coding for an olfactory receptor polypeptide, or a fragment or a variant thereof, is inserted in an appropriate vector, optionally after an appropriate cleavage of this amplified nucleic acid with one or several restriction endonucleases. The nucleic acid coding for an olfactory receptor polypeptide or a fragment or a variant thereof may be the resulting product of an amplification 25 reaction using a pair of primers according to the invention (by PCR, SDA, TAS, 3SR NASBA, TMA etc.).

C. ANTIBODIES THAT BIND OLFACTORY RECEPTOR POLYPEPTIDES

Any olfactory receptor polypeptide or whole protein may be used to generate antibodies capable of specifically binding to an expressed olfactory receptor protein or fragments thereof as 30 described.

One antibody composition of the invention is capable of specifically binding or specifically bind to the variant of the olfactory receptor protein of SEQ ID Nos 12-21. For an antibody composition to specifically bind to a first variant of olfactory receptor protein, it must demonstrate at least a 5%, 10%, 15%, 20%, 25%, 50%, or 100% greater binding affinity for a first variant of the 35 olfactory receptor protein than for a second variant of the olfactory receptor protein in an ELISA, RIA, or other antibody-based binding assay.

In a preferred embodiment, the invention concerns antibody compositions, either polyclonal or monoclonal, capable of selectively binding, or that selectively bind to an epitope-containing a polypeptide comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments". Preferred peptide fragments are portions of OLF1 to OLF10 polypeptides 5 that are located outside the transmembrane domains, most preferably peptide fragments naturally exposed on the cell membrane, particularly those that are available for binding to ligand molecules, either odorant substances or molecules or antibodies directed to the olfactory receptor polypeptides of the invention.

The invention also concerns a purified or isolated antibody capable of specifically binding to 10 a mutated olfactory receptor protein or to a fragment or variant thereof comprising an epitope of the mutated olfactory receptor protein. In another preferred embodiment, the present invention concerns an antibody capable of binding to a polypeptide comprising at least 10 consecutive amino acids of an olfactory receptor protein.

In a preferred embodiment, the invention concerns the use in the manufacture of antibodies 15 of a polypeptide comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments". Preferred peptide fragments are portions of OLF1 to OLF10 polypeptides that are located outside the transmembrane domains, most preferably peptide fragments naturally exposed on the cell membrane, particularly those that are available for recognition of ligand molecules, either odorant substances or molecules or antibodies directed to the olfactory 20 receptor polypeptides of the invention.

The olfactory receptor expressed from a DNA comprising at least one of the nucleic sequences of SEQ ID Nos 1-11 or a fragment or a variant thereof may also be used to generate antibodies capable of specifically binding to the expressed olfactory receptor or fragments or variants thereof. In a preferred embodiment, any of the polynucleotide fragment encoding a 25 polypeptide described in the section "Coding regions of the olfactory receptor gene" may be used to generate such antibodies.

Substantially pure protein or polypeptide is isolated from transfected or transformed cells containing an expression vector encoding the olfactory receptor protein or a portion thereof. The concentration of protein in the final preparation is adjusted, for example, by concentration on an 30 Amicon filter device, to the level of a few micrograms/ml. Monoclonal or polyclonal antibodies to the protein can then be prepared as follows:

1. Monoclonal Antibody Production by Hybridoma Fusion

Monoclonal antibody to epitopes in the olfactory receptor of the present invention or a portion thereof can be prepared from murine hybridomas according to the classical method of Kohler and 35 Milstein, (1975) or derivative methods thereof. Briefly, a mouse is repetitively inoculated with a few micrograms of the considered olfactory receptor or a portion thereof over a period of a few weeks. The mouse is then sacrificed, and the antibody producing cells of the spleen isolated. The spleen cells are

fused by means of polyethylene glycol with mouse myeloma cells, and the excess unfused cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). The successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate where growth of the culture is continued. Antibody-producing clones are identified by detection of 5 antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall, (1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Davis, L. et al.

2. Polyclonal Antibody Production by Immunization

10 Polyclonal antiserum containing antibodies to heterogeneous epitopes in the olfactory receptor of the present invention or a portion thereof can be prepared by immunizing suitable animals with the considered olfactory receptor or a portion thereof, which can be unmodified or modified to enhance immunogenicity. A suitable non-human animal, preferably a non-human mammal, is selected, usually a mouse, rat, rabbit, goat, or horse. Alternatively, a crude preparation which has been 15 enriched for olfactory receptor concentration can be used to generate antibodies. Such proteins, fragments or preparations are introduced into the non-human mammal in the presence of an appropriate adjuvant (e.g. aluminum hydroxide, RIBI, etc.) which is known in the art. In addition the protein, fragment or preparation can be pretreated with an agent which will increase antigenicity, such agents are known in the art and include, for example, methylated bovine serum albumin 20 (mBSA), bovine serum albumin (BSA), Hepatitis B surface antigen, and keyhole limpet hemocyanin (KLH). Serum from the immunized animal is collected, treated and tested according to known procedures. If the serum contains polyclonal antibodies to undesired epitopes, the polyclonal antibodies can be purified by immunoaffinity chromatography.

Effective polyclonal antibody production is affected by many factors related both to the antigen 25 and the host species. Also, host animals vary in response to site of inoculations and dose, with both inadequate or excessive doses of antigen resulting in low titer antisera. Small doses (ng level) of antigen administered at multiple intradermal sites appears to be most reliable. Techniques for producing and processing polyclonal antisera are known in the art, see for example, Mayer and Walker (1987). An effective immunization protocol for rabbits can be found in Vaitukaitis, J. et al. (1971).

30 Booster injections can be given at regular intervals, and antiserum harvested when antibody titer thereof, as determined semi-quantitatively, for example, by double immunodiffusion in agar against known concentrations of the antigen, begins to fall. See, for example, Ouchterlony et al., (1973). Plateau concentration of antibody is usually in the range of 0.1 to 0.2 mg/ml of serum. Affinity of the 35 antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher, (1980).

Antibody preparations prepared according to either the monoclonal or the polyclonal protocol are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances

in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample. The antibodies may also be used in therapeutic compositions for killing cells expressing the protein or reducing the levels of the protein in the body.

Non-human animals or mammals, whether wild-type or transgenic, which express a different 5 species of olfactory receptor than the one to which antibody binding is desired, and animals which do not express olfactory receptor (i.e. an olfactory receptor knock out animal as described herein) are particularly useful for preparing antibodies. Olfactory receptor knock out animals will recognize all or most of the exposed regions of an olfactory receptor protein as foreign antigens, and therefore produce antibodies with a wider array of olfactory receptor epitopes. Moreover, smaller 10 polypeptides with only 10 to 30 amino acids may be useful in obtaining specific binding to any one of the olfactory receptor proteins. In addition, the humoral immune system of animals which produce a species of olfactory receptor that resembles the antigenic sequence will preferentially recognize the differences between the animal's native olfactory receptor species and the antigen sequence, and produce antibodies to these unique sites in the antigen sequence. Such a technique 15 will be particularly useful in obtaining antibodies that specifically bind to any one of the olfactory receptor proteins.

The present invention also includes, chimeric single chain Fv antibody fragments (Martineau et al., 1998), antibody fragments obtained through phage display libraries (Ridder et al., 1995; Vaughan et al., 1995) and humanized antibodies (Reinmann et al., 1997; Leger et al., 1997).

20 The antibodies of the invention may be labeled by any one of the radioactive, fluorescent or enzymatic labels known in the art.

Consequently, the invention is also directed to a method for detecting specifically the presence of a polypeptide according to the invention in a biological sample, said method comprising the following steps :

25 a) bringing into contact the biological sample with an antibody according to the invention;

b) detecting the antigen-antibody complex formed.

Is also part of the invention a diagnostic kit for in vitro detecting the presence of a polypeptide according to the present invention in a biological sample, wherein said kit comprises:

30 a) a polyclonal or monoclonal antibody as described above, optionally labeled;

b) a reagent allowing the detection of the antigen-antibody complexes formed, said reagent carrying optionally a label, or being able to be recognized itself by a labeled reagent, more particularly in the case when the above-mentioned monoclonal or polyclonal antibody is not labeled by itself.

35 D. OLFACTORY RECEPTOR-RELATED BIALLELIC MARKERS

The invention also concerns olfactory receptor-related biallelic markers. As used herein the term "olfactory receptor-related biallelic marker" relates to a set of biallelic markers in linkage

disequilibrium with the olfactory receptor gene. The term olfactory receptor-related biallelic marker includes the biallelic markers designated A1 to A13.

The biallelic markers of the present invention, namely A1 to A13, are disclosed in Table 2 of Example 4. The 13 olfactory receptor-related biallelic markers, A1 to A13, are all located in the 5 genomic non coding regions of the olfactory gene cluster of the invention. Their precise location on the olfactory receptor genomic sequence and their single base polymorphism are indicated in Table 2 and also as features in the sequence listing for SEQ ID No 1. Appropriate pairs of primers allowing the amplification of a nucleic acid containing the polymorphic base of the disclosed olfactory receptor biallelic marker are also listed in Table 1 of Example 3 and in features of SEQ ID No 1.

10 In the present invention, the biallelic markers can be defined by nucleotide sequences corresponding to oligonucleotides of 47 bases in length comprising at the middle one of the polymorphic base. More particularly, the biallelic markers can be defined by the polynucleotides P1 to P13.

The biallelic markers contained in the olfactory gene cluster of the present invention, or a 15 busset of such biallelic markers, are useful tools to perform association studies, preferably to perform association studies between the statistically significant occurrence of an allele of said biallelic marker in the genome of an individual and a specific phenotype, including a phenotype consisting of an alteration of the olfactory perception of odorant substances or molecules by said individual. The biallelic markers of the invention can also be used, for example, in linkage analysis 20 in which evidence is sought for cosegregation between a locus and a putative trait locus using family studies, such as an alteration of olfactory perception. In addition, the biallelic markers of the invention may be included in the generation of any complete or partial genetic map of the human genome. These different uses are specifically contemplated in the present invention and claims.

1. Identification of biallelic markers

25 Any of a variety of methods can be used to screen a genomic fragment for single nucleotide polymorphisms such as differential hybridization with oligonucleotide probes, detection of changes in the mobility measured by gel electrophoresis or direct sequencing of the amplified nucleic acid. A preferred method for identifying biallelic markers involves comparative sequencing of genomic DNA fragments from an appropriate number of unrelated individuals.

30 In a first embodiment, DNA samples from unrelated individuals are pooled together, following which the genomic DNA of interest is amplified and sequenced. The nucleotide sequences thus obtained are then analyzed to identify significant polymorphisms. One of the major advantages of this method resides in the fact that the pooling of the DNA samples substantially reduces the number of DNA amplification reactions and sequencing reactions, which must be carried 35 out. Moreover, this method is sufficiently sensitive so that a biallelic marker obtained thereby usually shows a sufficient degree of informativeness to be useful in conducting association studies.

In a second embodiment, the DNA samples are not pooled and are therefore amplified and sequenced individually. This method is usually preferred when biallelic markers need to be identified in order to perform association studies within candidate genes. Preferably, highly relevant gene regions such as promoter regions or exon regions may be screened for biallelic markers. A 5 biallelic marker obtained using this method may show a lower degree of informativeness for conducting association studies, e.g. if the frequency of its less frequent allele may be less than about 10%. Such a biallelic marker will, however, be sufficiently informative to conduct association studies and it will further be appreciated that including less informative biallelic markers in the 10 genetic analysis studies of the present invention, may allow in some cases the direct identification of causal mutations, which may, depending on their penetrance, be rare mutations.

The following is a description of the various parameters of a preferred method used by the inventors for the identification of the biallelic markers of the present invention.

Genomic DNA Samples

The genomic DNA samples from which the biallelic markers of the present invention are 15 generated are preferably obtained from unrelated individuals corresponding to a heterogeneous population of known ethnic background. The number of individuals from whom DNA samples are obtained can vary substantially, preferably from about 10 to about 1000, preferably from about 50 to about 200 individuals. It is usually preferred to collect DNA samples from at least about 100 individuals in order to have sufficient polymorphic diversity in a given population to identify as 20 many markers as possible and to generate statistically significant results.

As for the source of the genomic DNA to be subjected to analysis, any test sample can be foreseen without any particular limitation. These test samples include biological samples, which can be tested by the methods of the present invention described herein, and include human and animal body fluids such as whole blood, serum, plasma, cerebrospinal fluid, urine, lymph fluids, and 25 various external secretions of the respiratory, intestinal and genitourinary tracts, tears, saliva, milk, white blood cells, myelomas and the like; biological fluids such as cell culture supernatants; fixed tissue specimens including tumor and non-tumor tissue and lymph node tissues; bone marrow aspirates and fixed cell specimens. The preferred source of genomic DNA used in the present invention is from peripheral venous blood of each donor. Techniques to prepare genomic DNA 30 from biological samples are well known to the skilled technician. Details of a preferred embodiment are provided in Example 2. The person skilled in the art can choose to amplify pooled or unpooled DNA samples.

DNA Amplification

The identification of biallelic markers in a sample of genomic DNA may be facilitated 35 through the use of DNA amplification methods. DNA samples can be pooled or unpooled for the amplification step. DNA amplification techniques are well known to those skilled in the art.

Amplification techniques that can be used in the context of the present invention include, but are not limited to, the ligase chain reaction (LCR) described in EP-A- 320 308, WO 9320227 and EP-A-439 182, the polymerase chain reaction (PCR, RT-PCR) and techniques such as the nucleic acid sequence based amplification (NASBA) described in Guatelli J.C., et al.(1990) and in Compton 5 J.(1991), Q-beta amplification as described in European Patent Application No 4544610, strand displacement amplification as described in Walker et al.(1996) and EP A 684 315 and, target mediated amplification as described in PCT Publication WO 9322461.

LCR and Gap LCR are exponential amplification techniques, both depend on DNA ligase to join adjacent primers annealed to a DNA molecule. In Ligase Chain Reaction (LCR), probe pairs 10 are used which include two primary (first and second) and two secondary (third and fourth) probes, all of which are employed in molar excess to target. The first probe hybridizes to a first segment of the target strand and the second probe hybridizes to a second segment of the target strand, the first and second segments being contiguous so that the primary probes abut one another in 5' phosphate- 3'hydroxyl relationship, and so that a ligase can covalently fuse or ligate the two probes into a fused 15 product. In addition, a third (secondary) probe can hybridize to a portion of the first probe and a fourth (secondary) probe can hybridize to a portion of the second probe in a similar abutting fashion. Of course, if the target is initially double stranded, the secondary probes also will hybridize to the target complement in the first instance. Once the ligated strand of primary probes is separated from 20 the target strand, it will hybridize with the third and fourth probes, which can be ligated to form a complementary, secondary ligated product. It is important to realize that the ligated products are functionally equivalent to either the target or its complement. By repeated cycles of hybridization and ligation, amplification of the target sequence is achieved. A method for multiplex LCR has also been described (WO 9320227). Gap LCR (GLCR) is a version of LCR where the probes are not 25 adjacent but are separated by 2 to 3 bases.

25 For amplification of mRNAs, it is within the scope of the present invention to reverse transcribe mRNA into cDNA followed by polymerase chain reaction (RT-PCR); or, to use a single enzyme for both steps as described in U.S. Patent No. 5,322,770 or, to use Asymmetric Gap LCR (RT-AGLCR) as described by Marshall et al.(1994). AGLCR is a modification of GLCR that allows the amplification of RNA.

30 The PCR technology is the preferred amplification technique used in the present invention. A variety of PCR techniques are familiar to those skilled in the art. For a review of PCR technology, see White (1997) and the publication entitled "PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press). In each of these PCR procedures, PCR primers on either side of the nucleic acid sequences to be amplified are added to a suitably prepared nucleic acid sample along 35 with dNTPs and a thermostable polymerase such as Taq polymerase, Pfu polymerase, or Vent polymerase. The nucleic acid in the sample is denatured and the PCR primers are specifically hybridized to complementary nucleic acid sequences in the sample. The hybridized primers are

extended. Thereafter, another cycle of denaturation, hybridization, and extension is initiated. The cycles are repeated multiple times to produce an amplified fragment containing the nucleic acid sequence between the primer sites. PCR has further been described in several patents including US Patents 4,683,195; 4,683,202; and 4,965,188.

5 The PCR technology is the preferred amplification technique used to identify new biallelic markers. A typical example of a PCR reaction suitable for the purposes of the present invention is provided in Example 3.

One of the aspects of the present invention is a method for the amplification of the human olfactory receptor gene, particularly of a fragment of the genomic sequence of SEQ ID No 1 or of 10 the coding region sequences of SEQ ID Nos 2-11, or a fragment or a variant thereof in a test sample, preferably using the PCR technology. This method comprises the steps of:

- 15 a) contacting a test sample with amplification reaction reagents comprising a pair of amplification primers as described above and located on either side of the polynucleotide region to be amplified, and
- b) optionally, detecting the amplification products.

The invention also concerns a kit for the amplification of an olfactory receptor gene sequence, particularly of a portion of the genomic sequence of SEQ ID No 1 or of the coding region sequences of SEQ ID Nos 2-11, or a variant thereof in a test sample, wherein said kit comprises:

- 20 a) a pair of oligonucleotide primers located on either side of the olfactory receptor region to be amplified;
- b) optionally, the reagents necessary for performing the amplification reaction.

In one embodiment of the above amplification method and kit, the amplification product is detected by hybridization with a labeled probe having a sequence which is complementary to the amplified region. In another embodiment of the above amplification method and kit, primers 25 comprise a sequence which is selected from the group consisting of the nucleotide sequences of B1 to B11, C1 to C11, D1 to D13, and E1 to E13.

In a first embodiment of the present invention, biallelic markers are identified using genomic sequence information generated by the inventors. Sequenced genomic DNA fragments are used to design primers for the amplification of 500 bp fragments. These 500 bp fragments are amplified 30 from genomic DNA and are scanned for biallelic markers. Primers may be designed using the OSP software (Hillier L. and Green P., 1991). All primers may contain, upstream of the specific target bases, a common oligonucleotide tail that serves as a sequencing primer. Those skilled in the art are familiar with primer extensions, which can be used for these purposes.

Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms

35 The amplification products generated as described above, are then sequenced using any method known and available to the skilled technician. Methods for sequencing DNA using either the dideoxy-mediated method (Sanger method) or the Maxam-Gilbert method are widely known to

those of ordinary skill in the art. Such methods are for example disclosed in Sambrook et al.(1989). Alternative approaches include hybridization to high-density DNA probe arrays as described in Chee et al.(1996).

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Following gel image analysis and DNA sequence extraction, sequence data are automatically processed with adequate software to assess sequence quality.

A polymorphism analysis software is used that detects the presence of biallelic sites among individual or pooled amplified fragment sequences. Polymorphism search is based on the presence of superimposed peaks in the electrophoresis pattern. These peaks which present distinct colors correspond to two different nucleotides at the same position on the sequence. The polymorphism has to be detected on both strands for validation.

Validation Of The Biallelic Markers Of The Present Invention

The polymorphisms are evaluated for their usefulness as genetic markers by validating that both alleles are present in a population. Validation of the biallelic markers is accomplished by genotyping a group of individuals by a method of the invention and demonstrating that both alleles are present. Microsequencing is a preferred method of genotyping alleles. The validation by genotyping step may be performed on individual samples derived from each individual in the group or by genotyping a pooled sample derived from more than one individual. The group can be as small as one individual if that individual is heterozygous for the allele in question. Preferably the group contains at least three individuals, more preferably the group contains five or six individuals, so that a single validation test will be more likely to result in the validation of more of the biallelic markers that are being tested. It should be noted, however, that when the validation test is performed on a small group it may result in a false negative result if as a result of sampling error none of the individuals tested carries one of the two alleles. Thus, the validation process is less useful in demonstrating that a particular initial result is an artifact, than it is at demonstrating that there is a *bona fide* biallelic marker at a particular position in a sequence. All of the genotyping, haplotyping, association, and interaction study methods of the invention may optionally be performed solely with validated biallelic markers.

30 2. Genotyping of biallelic markers

The polymorphisms identified above can be further confirmed and their respective frequencies can be determined through various methods using the previously described primers and probes. These methods can also be useful for genotyping either new populations in association studies or individuals in the context of detection of alleles of biallelic markers which are known to be associated with a given trait. Those skilled in the art should note that the methods described below can be equally performed on individual or pooled DNA samples.

Once a given polymorphic site has been found and characterized as a biallelic marker as described above, several methods can be used in order to determine the specific allele carried by an individual at the given polymorphic base.

The identification of biallelic markers described previously allows the design of appropriate 5 primers to amplify a region of the olfactory receptor gene cluster containing the polymorphic site of interest and for the detection of such polymorphisms.

Genotyping can be performed using similar methods as those described above for the identification of the biallelic markers, or using other genotyping methods such as those further described below. In preferred embodiments, the comparison of sequences of amplified genomic 10 fragments from different individuals is used to identify new biallelic markers whereas microsequencing is used for genotyping known biallelic markers in diagnostic and genetic analysis applications.

In one embodiment the invention encompasses methods of genotyping comprising determining the identity of a nucleotide at an olfactory receptor-related biallelic marker or the 15 complement thereof in a biological sample; optionally, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof; optionally, wherein said biological sample is derived from a single subject; optionally, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome; optionally, wherein said biological sample is derived 20 from multiple subjects; Optionally, the genotyping methods of the invention encompass methods with any further limitation described in this disclosure, or those following, specified alone or in any combination; Optionally, said method is performed *in vitro*; optionally, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step; Optionally, wherein said amplifying is performed by PCR, LCR, or replication of a recombinant 25 vector comprising an origin of replication and said fragment in a host cell; optionally, wherein said determining is performed by a hybridization assay, a sequencing assay, a microsequencing assay, or an enzyme-based mismatch detection assay.

Source of Nucleic Acids for genotyping

Any source of nucleic acids, in purified or non-purified form, can be utilized as the starting 30 nucleic acid, provided it contains or is suspected of containing the specific nucleic acid sequence desired. DNA or RNA may be extracted from cells, tissues, body fluids and the like as described above. While nucleic acids for use in the genotyping methods of the invention can be derived from any mammalian source, the test subjects and individuals from which nucleic acid samples are taken are generally understood to be human.

Amplification Of DNA Fragments Comprising Biallelic Markers

Methods and polynucleotides are provided to amplify a segment of nucleotides comprising one or more biallelic marker of the present invention. It will be appreciated that amplification of DNA fragments comprising biallelic markers may be used in various methods and for various purposes and is not restricted to genotyping. Nevertheless, many genotyping methods, although not all, require the previous amplification of the DNA region carrying the biallelic marker of interest. Such methods specifically increase the concentration or total number of sequences that span the biallelic marker or include that site and sequences located either distal or proximal to it. Diagnostic assays may also rely on amplification of DNA segments carrying a biallelic marker of the present invention. Amplification of DNA may be achieved by any method known in the art. Amplification techniques are described above in the section entitled, "DNA amplification."

Some of these amplification methods are particularly suited for the detection of single nucleotide polymorphisms and allow the simultaneous amplification of a target sequence and the identification of the polymorphic nucleotide as it is further described below.

15 The identification of biallelic markers as described above allows the design of appropriate oligonucleotides, which can be used as primers to amplify DNA fragments comprising the biallelic markers of the present invention. Amplification can be performed using the primers initially used to discover new biallelic markers which are described herein or any set of primers allowing the amplification of a DNA fragment comprising a biallelic marker of the present invention.

20 In some embodiments the present invention provides primers for amplifying a DNA fragment containing one or more biallelic markers of the present invention. Preferred amplification primers are listed in Example 3. It will be appreciated that the primers listed are merely exemplary and that any other set of primers which produce amplification products containing one or more biallelic markers of the present invention are also of use.

25 The spacing of the primers determines the length of the segment to be amplified. In the context of the present invention, amplified segments carrying biallelic markers can range in size from at least about 25 bp to 35 kbp. Amplification fragments from 25-3000 bp are typical, fragments from 50-1000 bp are preferred and fragments from 100-600 bp are highly preferred. It will be appreciated that amplification primers for the biallelic markers may be any sequence which 30 allow the specific amplification of any DNA fragment carrying the markers. Amplification primers may be labeled or immobilized on a solid support as described in "Oligonucleotide probes and primers".

Methods of Genotyping DNA samples for Biallelic Markers

Any method known in the art can be used to identify the nucleotide present at a biallelic marker site. Since the biallelic marker allele to be detected has been identified and specified in the present invention, detection will prove simple for one of ordinary skill in the art by employing any of a number of techniques. Many genotyping methods require the previous amplification of the

DNA region carrying the biallelic marker of interest. While the amplification of target or signal is often preferred at present, ultrasensitive detection methods which do not require amplification are also encompassed by the present genotyping methods. Methods well-known to those skilled in the art that can be used to detect biallelic polymorphisms include methods such as, conventional dot blot 5 analyzes, single strand conformational polymorphism analysis (SSCP) described by Orita et al.(1989), denaturing gradient gel electrophoresis (DGGE), heteroduplex analysis, mismatch cleavage detection, and other conventional techniques as described in Sheffield et al.(1991), White et al.(1992), Grompe et al.(1989 and 1993). Another method for determining the identity of the 10 nucleotide present at a particular polymorphic site employs a specialized exonuclease-resistant nucleotide derivative as described in US patent 4,656,127.

Preferred methods involve directly determining the identity of the nucleotide present at a biallelic marker site by sequencing assay, enzyme-based mismatch detection assay, or hybridization assay. The following is a description of some preferred methods. A highly preferred method is the 15 microsequencing technique. The term "sequencing" is generally used herein to refer to polymerase extension of duplex primer/template complexes and includes both traditional sequencing and microsequencing.

1) Sequencing Assays

The nucleotide present at a polymorphic site can be determined by sequencing methods. In a preferred embodiment, DNA samples are subjected to PCR amplification before sequencing as 20 described above. DNA sequencing methods are described in "Sequencing Of Amplified Genomic DNA And Identification Of Single Nucleotide Polymorphisms".

Preferably, the amplified DNA is subjected to automated dideoxy terminator sequencing reactions using a dye-primer cycle sequencing protocol. Sequence analysis allows the identification of the base present at the biallelic marker site.

25 2) Microsequencing Assays

In microsequencing methods, the nucleotide at a polymorphic site in a target DNA is detected by a single nucleotide primer extension reaction. This method involves appropriate microsequencing primers which, hybridize just upstream of the polymorphic base of interest in the target nucleic acid. A polymerase is used to specifically extend the 3' end of the primer with one 30 single ddNTP (chain terminator) complementary to the nucleotide at the polymorphic site. Next the identity of the incorporated nucleotide is determined in any suitable way.

Typically, microsequencing reactions are carried out using fluorescent ddNTPs and the extended microsequencing primers are analyzed by electrophoresis on ABI 377 sequencing machines to determine the identity of the incorporated nucleotide as described in EP 412 883.

35 Alternatively capillary electrophoresis can be used in order to process a higher number of assays simultaneously. An example of a typical microsequencing procedure that can be used in the context of the present invention is provided in Example 5.

Different approaches can be used for the labeling and detection of ddNTPs. A homogeneous phase detection method based on fluorescence resonance energy transfer has been described by Chen and Kwok (1997) and Chen et al.(1997). In this method, amplified genomic DNA fragments containing polymorphic sites are incubated with a 5'-fluorescein-labeled primer in the presence of 5 allelic dye-labeled dideoxyribonucleoside triphosphates and a modified Taq polymerase. The dye-labeled primer is extended one base by the dye-terminator specific for the allele present on the template. At the end of the genotyping reaction, the fluorescence intensities of the two dyes in the reaction mixture are analyzed directly without separation or purification. All these steps can be performed in the same tube and the fluorescence changes can be monitored in real time.

10 Alternatively, the extended primer may be analyzed by MALDI-TOF Mass Spectrometry. The base at the polymorphic site is identified by the mass added onto the microsequencing primer (see Haff and Smirnov, 1997).

Microsequencing may be achieved by the established microsequencing method or by developments or derivatives thereof. Alternative methods include several solid-phase 15 microsequencing techniques. The basic microsequencing protocol is the same as described previously, except that the method is conducted as a heterogeneous phase assay, in which the primer or the target molecule is immobilized or captured onto a solid support. To simplify the primer separation and the terminal nucleotide addition analysis, oligonucleotides are attached to solid supports or are modified in such ways that permit affinity separation as well as polymerase 20 extension. The 5' ends and internal nucleotides of synthetic oligonucleotides can be modified in a number of different ways to permit different affinity separation approaches, e.g., biotinylation. If a single affinity group is used on the oligonucleotides, the oligonucleotides can be separated from the incorporated terminator reagent. This eliminates the need of physical or size separation. More than one oligonucleotide can be separated from the terminator reagent and analyzed simultaneously if 25 more than one affinity group is used. This permits the analysis of several nucleic acid species or more nucleic acid sequence information per extension reaction. The affinity group need not be on the priming oligonucleotide but could alternatively be present on the template. For example, immobilization can be carried out via an interaction between biotinylated DNA and streptavidin-coated microtitration wells or avidin-coated polystyrene particles. In the same manner, 30 oligonucleotides or templates may be attached to a solid support in a high-density format. In such solid phase microsequencing reactions, incorporated ddNTPs can be radiolabeled (Syvänen, 1994) or linked to fluorescein (Livak and Hainer, 1994). The detection of radiolabeled ddNTPs can be achieved through scintillation-based techniques. The detection of fluorescein-linked ddNTPs can be based on the binding of antifluorescein antibody conjugated with alkaline phosphatase, followed by 35 incubation with a chromogenic substrate (such as *p*-nitrophenyl phosphate). Other possible reporter-detection pairs include: ddNTP linked to dinitrophenyl (DNP) and anti-DNP alkaline phosphatase conjugate (Harju et al., 1993) or biotinylated ddNTP and horseradish peroxidase-conjugated

streptavidin with *o*-phenylenediamine as a substrate (WO 92/15712). As yet another alternative solid-phase microsequencing procedure, Nyren et al.(1993) described a method relying on the detection of DNA polymerase activity by an enzymatic luminometric inorganic pyrophosphate detection assay (ELIDA).

5 Pastinen et al.(1997) describe a method for multiplex detection of single nucleotide polymorphism in which the solid phase minisequencing principle is applied to an oligonucleotide array format. High-density arrays of DNA probes attached to a solid support (DNA chips) are further described below.

In one aspect the present invention provides polynucleotides and methods to genotype one or 10 more biallelic markers of the present invention by performing a microsequencing assay. Preferred microsequencing primers include the nucleotide sequences D1 to Dn and E1 to En. It will be appreciated that the microsequencing primers listed in Example 5 are merely exemplary and that, any primer having a 3' end immediately adjacent to the polymorphic nucleotide may be used. Similarly, it will be appreciated that microsequencing analysis may be performed for any biallelic 15 marker or any combination of biallelic markers of the present invention. One aspect of the present invention is a solid support which includes one or more microsequencing primers listed in Example 5, or fragments comprising at least 8, 12, 15, 20, 25, 30, 40, or 50 consecutive nucleotides thereof, to the extent that such lengths are consistent with the primer described, and having a 3' terminus immediately upstream of the corresponding biallelic marker, for determining the identity of a 20 nucleotide at a biallelic marker site.

3) Mismatch detection assays based on polymerases and ligases

In one aspect the present invention provides polynucleotides and methods to determine the allele of one or more biallelic markers of the present invention in a biological sample, by mismatch detection assays based on polymerases and/or ligases. These assays are based on the specificity of 25 polymerases and ligases. Polymerization reactions places particularly stringent requirements on correct base pairing of the 3' end of the amplification primer and the joining of two oligonucleotides hybridized to a target DNA sequence is quite sensitive to mismatches close to the ligation site, especially at the 3' end. Methods, primers and various parameters to amplify DNA fragments comprising biallelic markers of the present invention are further described above in "Amplification 30 Of DNA Fragments Comprising Biallelic Markers".

Allele Specific Amplification Primers

Discrimination between the two alleles of a biallelic marker can also be achieved by allele specific amplification, a selective strategy, whereby one of the alleles is amplified without amplification of the other allele. For allele specific amplification, at least one member of the pair of 35 primers is sufficiently complementary with a region of an olfactory receptor gene comprising the polymorphic base of a biallelic marker of the present invention to hybridize therewith and to initiate

the amplification. Such primers are able to discriminate between the two alleles of a biallelic marker.

This is accomplished by placing the polymorphic base at the 3' end of one of the amplification primers. Because the extension forms from the 3' end of the primer, a mismatch at or 5 near this position has an inhibitory effect on amplification. Therefore, under appropriate amplification conditions, these primers only direct amplification on their complementary allele. Determining the precise location of the mismatch and the corresponding assay conditions are well within the ordinary skill in the art.

Ligation/Amplification Based Methods

10 The "Oligonucleotide Ligation Assay" (OLA) uses two oligonucleotides which are designed to be capable of hybridizing to abutting sequences of a single strand of a target molecules. One of the oligonucleotides is biotinylated, and the other is detectably labeled. If the precise complementary sequence is found in a target molecule, the oligonucleotides will hybridize such that their termini abut, and create a ligation substrate that can be captured and detected. OLA is capable 15 of detecting single nucleotide polymorphisms and may be advantageously combined with PCR as described by Nickerson et al.(1990). In this method, PCR is used to achieve the exponential amplification of target DNA, which is then detected using OLA.

Other amplification methods which are particularly suited for the detection of single nucleotide polymorphism include LCR (ligase chain reaction), Gap LCR (GLCR) which are 20 described above in "DNA Amplification". LCR uses two pairs of probes to exponentially amplify a specific target. The sequences of each pair of oligonucleotides, is selected to permit the pair to hybridize to abutting sequences of the same strand of the target. Such hybridization forms a substrate for a template-dependant ligase. In accordance with the present invention, LCR can be performed with oligonucleotides having the proximal and distal sequences of the same strand of a 25 biallelic marker site. In one embodiment, either oligonucleotide will be designed to include the biallelic marker site. In such an embodiment, the reaction conditions are selected such that the oligonucleotides can be ligated together only if the target molecule either contains or lacks the specific nucleotide that is complementary to the biallelic marker on the oligonucleotide. In an alternative embodiment, the oligonucleotides will not include the biallelic marker, such that when 30 they hybridize to the target molecule, a "gap" is created as described in WO 90/01069. This gap is then "filled" with complementary dNTPs (as mediated by DNA polymerase), or by an additional pair of oligonucleotides. Thus at the end of each cycle, each single strand has a complement capable of serving as a target during the next cycle and exponential allele-specific amplification of the desired sequence is obtained.

35 Ligase/Polymerase-mediated Genetic Bit AnalysisTM is another method for determining the identity of a nucleotide at a preselected site in a nucleic acid molecule (WO 95/21271). This method involves the incorporation of a nucleoside triphosphate that is complementary to the nucleotide

present at the preselected site onto the terminus of a primer molecule, and their subsequent ligation to a second oligonucleotide. The reaction is monitored by detecting a specific label attached to the reaction's solid phase or by detection in solution.

4) Hybridization Assay Methods

5 A preferred method of determining the identity of the nucleotide present at a biallelic marker site involves nucleic acid hybridization. The hybridization probes, which can be conveniently used in such reactions, preferably include the probes defined herein. Any hybridization assay may be used including Southern hybridization, Northern hybridization, dot blot hybridization and solid-phase hybridization (see Sambrook et al., 1989).

10 Hybridization refers to the formation of a duplex structure by two single stranded nucleic acids due to complementary base pairing. Hybridization can occur between exactly complementary nucleic acid strands or between nucleic acid strands that contain minor regions of mismatch. Specific probes can be designed that hybridize to one form of a biallelic marker and not to the other and therefore are able to discriminate between different allelic forms. Allele-specific probes are

15 often used in pairs, one member of a pair showing perfect match to a target sequence containing the original allele and the other showing a perfect match to the target sequence containing the alternative allele. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. Stringent, sequence specific hybridization conditions,

20 under which a probe will hybridize only to the exactly complementary target sequence are well known in the art (Sambrook et al., 1989). Stringent conditions are sequence dependent and will be different in different circumstances. Generally, stringent conditions are selected to be about 5°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. Although such hybridization can be performed in solution, it is preferred to employ a solid-

25 phase hybridization assay. The target DNA comprising a biallelic marker of the present invention may be amplified prior to the hybridization reaction. The presence of a specific allele in the sample is determined by detecting the presence or the absence of stable hybrid duplexes formed between the probe and the target DNA. The detection of hybrid duplexes can be carried out by a number of methods. Various detection assay formats are well known which utilize detectable labels bound to

30 either the target or the probe to enable detection of the hybrid duplexes. Typically, hybridization duplexes are separated from unhybridized nucleic acids and the labels bound to the duplexes are then detected. Those skilled in the art will recognize that wash steps may be employed to wash away excess target DNA or probe as well as unbound conjugate. Further, standard heterogeneous assay formats are suitable for detecting the hybrids using the labels present on the primers and probes.

35 Two recently developed assays allow hybridization-based allele discrimination with no need for separations or washes (see Landegren U. et al., 1998). The TaqMan assay takes advantage of the 5' nuclease activity of Taq DNA polymerase to digest a DNA probe annealed specifically to the

accumulating amplification product. TaqMan probes are labeled with a donor-acceptor dye pair that interacts via fluorescence energy transfer. Cleavage of the TaqMan probe by the advancing polymerase during amplification dissociates the donor dye from the quenching acceptor dye, greatly increasing the donor fluorescence. All reagents necessary to detect two allelic variants can be 5 assembled at the beginning of the reaction and the results are monitored in real time (see Livak et al., 1995). In an alternative homogeneous hybridization based procedure, molecular beacons are used for allele discriminations. Molecular beacons are hairpin-shaped oligonucleotide probes that report the presence of specific nucleic acids in homogeneous solutions. When they bind to their targets they undergo a conformational reorganization that restores the fluorescence of an internally 10 quenched fluorophore (Tyagi et al., 1998).

The polynucleotides provided herein can be used to produce probes which can be used in hybridization assays for the detection of biallelic marker alleles in biological samples. These probes are characterized in that they preferably comprise between 8 and 50 nucleotides, and in that they are sufficiently complementary to a sequence comprising a biallelic marker of the present invention to 15 hybridize thereto and preferably sufficiently specific to be able to discriminate the targeted sequence for only one nucleotide variation. A particularly preferred probe is 25 nucleotides in length. Preferably the biallelic marker is within 4 nucleotides of the center of the polynucleotide probe. In particularly preferred probes, the biallelic marker is at the center of said polynucleotide. Preferred probes comprise a nucleotide sequence selected from the group consisting of amplicons listed in 20 Table 1 and the sequences complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. Preferred probes comprise a nucleotide sequence selected from the group consisting of P1 to P13 and the sequences complementary thereto. In preferred embodiments the polymorphic base(s) are within 5, 4, 3, 2, 1, nucleotides of the center 25 of the said polynucleotide, more preferably at the center of said polynucleotide.

Preferably the probes of the present invention are labeled or immobilized on a solid support. Labels and solid supports are further described in "Oligonucleotide Probes and Primers". The probes can be non-extendable as described in "Oligonucleotide Probes and Primers".

By assaying the hybridization to an allele specific probe, one can detect the presence or 30 absence of a biallelic marker allele in a given sample. High-Throughput parallel hybridization in array format is specifically encompassed within "hybridization assays" and are described below.

5) Hybridization To Addressable Arrays Of Oligonucleotides

Hybridization assays based on oligonucleotide arrays rely on the differences in hybridization 35 stability of short oligonucleotides to perfectly matched and mismatched target sequence variants. Efficient access to polymorphism information is obtained through a basic structure comprising high-density arrays of oligonucleotide probes attached to a solid support (e.g., the chip) at selected

positions. Each DNA chip can contain thousands to millions of individual synthetic DNA probes arranged in a grid-like pattern and miniaturized to the size of a dime.

The chip technology has already been applied with success in numerous cases. For example, the screening of mutations has been undertaken in the BRCA1 gene, in *S. cerevisiae* mutant strains, 5 and in the protease gene of HIV-1 virus (Hacia et al., 1996; Shoemaker et al., 1996; Kozal et al., 1996). Chips of various formats for use in detecting biallelic polymorphisms can be produced on a customized basis by Affymetrix (GeneChip™), Hyseq (HyChip and HyGnostics), and Protogene Laboratories.

In general, these methods employ arrays of oligonucleotide probes that are complementary 10 to target nucleic acid sequence segments from an individual which, target sequences include a polymorphic marker. EP 785280 describes a tiling strategy for the detection of single nucleotide polymorphisms. Briefly, arrays may generally be “tiled” for a large number of specific polymorphisms. By “tiling” is generally meant the synthesis of a defined set of oligonucleotide probes which is made up of a sequence complementary to the target sequence of interest, as well as 15 preselected variations of that sequence, e.g., substitution of one or more given positions with one or more members of the basis set of nucleotides. Tiling strategies are further described in PCT application No. WO 95/11995. In a particular aspect, arrays are tiled for a number of specific, identified biallelic marker sequences. In particular, the array is tiled to include a number of detection blocks, each detection block being specific for a specific biallelic marker or a set of 20 biallelic markers. For example, a detection block may be tiled to include a number of probes, which span the sequence segment that includes a specific polymorphism. To ensure probes that are complementary to each allele, the probes are synthesized in pairs differing at the biallelic marker. In addition to the probes differing at the polymorphic base, monosubstituted probes are also generally tiled within the detection block. These monosubstituted probes have bases at and up to a certain 25 number of bases in either direction from the polymorphism, substituted with the remaining nucleotides (selected from A, T, G, C and U). Typically the probes in a tiled detection block will include substitutions of the sequence positions up to and including those that are 5 bases away from the biallelic marker. The monosubstituted probes provide internal controls for the tiled array, to distinguish actual hybridization from artefactual cross-hybridization. Upon completion of 30 hybridization with the target sequence and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data from the scanned array is then analyzed to identify which allele or alleles of the biallelic marker are present in the sample. Hybridization and scanning may be carried out as described in PCT application No. WO 92/10092 and WO 95/11995 and US patent No. 5,424,186.

35 Thus, in some embodiments, the chips may comprise an array of nucleic acid sequences of fragments of about 15 nucleotides in length. In further embodiments, the chip may comprise an array including at least one of the sequences selected from the group consisting of amplicons listed

in table 1 and the sequences complementary thereto, or a fragment thereof, said fragment comprising at least about 8 consecutive nucleotides, preferably 10, 15, 20, more preferably 25, 30, 40, 47, or 50 consecutive nucleotides and containing a polymorphic base. In preferred embodiments the polymorphic base is within 5, 4, 3, 2, 1, nucleotides of the center of the said polynucleotide, more 5 preferably at the center of said polynucleotide. In some embodiments, the chip may comprise an array of at least 2, 3, 4, 5, 6, 7, 8 or more of these polynucleotides of the invention. Solid supports and polynucleotides of the present invention attached to solid supports are further described in "Oligonucleotide Probes And Primers".

6) Integrated Systems

10 Another technique, which may be used to analyze polymorphisms, includes multicomponent integrated systems, which miniaturize and compartmentalize processes such as PCR and capillary electrophoresis reactions in a single functional device. An example of such technique is disclosed in US patent 5,589,136 which describes the integration of PCR amplification and capillary electrophoresis in chips.

15 Integrated systems can be envisaged mainly when microfluidic systems are used. These systems comprise a pattern of microchannels designed onto a glass, silicon, quartz, or plastic wafer included on a microchip. The movements of the samples are controlled by electric, electroosmotic or hydrostatic forces applied across different areas of the microchip to create functional microscopic valves and pumps with no moving parts.

20 For genotyping biallelic markers, the microfluidic system may integrate nucleic acid amplification, microsequencing, capillary electrophoresis and a detection method such as laser-induced fluorescence detection.

E. EXPRESSION OF AN OL1 TO OLF10 CODING POLYNUCLEOTIDE

Any of the coding polynucleotides of the invention may be inserted into recombinant vectors 25 for expression in a recombinant host cell or a recombinant host organism.

Thus, the present invention also encompasses a family of recombinant vectors that contains a coding polynucleotide from the group of coding polynucleotides OLF1 to OLF10 genes. Consequently, the present invention further deals with a recombinant vector comprising a polynucleotide comprising any of the coding sequence of SEQ ID No 1, preferably those selected 30 from the group consisting of SEQ ID Nos 2-11.

In a first preferred embodiment, the present invention relates to expression vectors which include nucleic acids encoding an olfactory receptor protein described herein under the control of an exogenous regulatory sequence.

In a second preferred embodiment, a recombinant vector of the invention is used to amplify 35 the inserted polynucleotide derived from an olfactory receptor genomic sequence selected from the group consisting of the nucleic acids of SEQ ID No 1 and of olfactory receptor cDNAs, for example

the open reading frames of SEQ ID Nos 2-11, in a suitable cell host, this polynucleotide being amplified at every time that the recombinant vector replicates.

More particularly, the present invention relates to expression vectors which include nucleic acids encoding an olfactory receptor protein, preferably the olfactory receptor proteins of the amino acid sequence of SEQ ID Nos 12-21 or variants or fragments thereof, under the control of an exogenous regulatory sequence.

Generally, a recombinant vector of the invention may comprise any of the polynucleotides described herein, including regulatory sequences, and coding sequences, as well as any olfactory receptor primer or probe as defined above. More particularly, the recombinant vectors of the present invention can comprise any of the polynucleotides described in the "Coding Regions of the olfactory receptor gene" section, "Genomic sequence of the olfactory receptor gene" section, the "Oligonucleotide Probes And Primers" section and the "Polynucleotide constructs" section.

Some of the elements which can be found in the vectors of the present invention are described in further detail in the following sections.

15 **Vectors**

A recombinant vector according to the invention comprises, but is not limited to, a YAC (Yeast Artificial Chromosome), a BAC (Bacterial Artificial Chromosome), a phage, a phagemid, a cosmid, a plasmid or even a linear DNA molecule which may consist of a chromosomal, non-chromosomal and synthetic DNA. Such a recombinant vector can comprise a transcriptional unit comprising an assembly of

(1) a genetic element or elements having a regulatory role in gene expression, for example promoters or enhancers. Enhancers are cis-acting elements of DNA, usually from about 10 to 300 bp that act on the promoter to increase the transcription.

(2) a structural or coding sequence which is transcribed into mRNA and eventually 25 translated into a polypeptide, and

(3) appropriate transcription initiation and termination sequences. Structural units intended for use in yeast or eukaryotic expression systems preferably include a leader sequence enabling extracellular secretion of translated protein by a host cell. Alternatively, where recombinant protein is expressed without a leader or transport sequence, it may include an N-terminal residue. This residue may or may not be subsequently cleaved from the expressed 30 recombinant protein to provide a final product.

Generally, recombinant expression vectors will include origins of replication, selectable markers permitting transformation of the host cell, and a promoter derived from a highly expressed gene to direct transcription of a downstream structural sequence. The heterologous structural sequence is assembled in appropriate phase with translation initiation and termination sequences, and preferably a leader sequence capable of directing secretion of translated protein into the 35 periplasmic space or extracellular medium.

The selectable marker genes for selection of transformed host cells are preferably dihydrofolate reductase or neomycin resistance for eukaryotic cell culture, TRP1 for *S. cerevisiae* or tetracycline, rifampicine or ampicillin resistance in *E. coli*, or levan saccharase for mycobacteria.

For facilitating the purification of the expressed protein and increasing its stability, the 5 coding sequence of an olfactory receptor according to the invention can be fused in its N- or C- terminus with protein such as MBP (maltose binding protein) and GST (Glutathione S transferase) or with tag such as poly-histidine tag, Strep tag, Bio tag, and flag peptide epitope tag, those being detailed below. Thioredoxin can be eventually inserted between the olfactory receptor and the tag.

Useful expression vectors for bacterial use are constructed by inserting a structural DNA 10 sequence encoding a desired polypeptide with suitable translation initiation and termination signals in operable reading phase with a functional promoter. The vector will comprise one or more phenotypic selectable markers and an origin of replication to ensure maintenance of the vector and to, if desirable, provide amplification within the host.

As a representative but non-limiting example, useful expression vectors for bacterial use can 15 comprise a selectable marker and bacterial origin of replication derived from commercially available plasmids comprising genetic elements of pBR322 (ATCC 37017). Such commercial vectors include, for example, pKK223-3 (Pharmacia, Uppsala, Sweden), and GEM1 (Promega Biotec, Madison, WI, USA).

Large numbers of suitable vectors and promoters are known to those of skill in the art, and 20 commercially available, such as bacterial vectors : pQE70, pQE60, pQE-9 (Qiagen), pbs, pD10, phagescript, psiX174, pbluescript SK, pbsks, pNH8A, pNH16A, pNH18A, pNH46A (Stratagene); ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 (Pharmacia); or eukaryotic vectors : pWLNEO, pSV2CAT, pOG44, pXT1, pSG (Stratagene); pSVK3, pBPV, pMSG, pSVL (Pharmacia); baculovirus transfer vector pVL1392/1393 (Pharmingen); pQE-30 (QIAexpress).

25 A suitable vector for the expression of the olfactory receptor above-defined or their peptide fragments is baculovirus vector that can be propagated in insect cells and in insect cell lines. A specific suitable host vector system is the pVL1392/1393 baculovirus transfer vector (Pharmingen) that is used to transfect the SF9 cell line (ATCC N°CRL 1711) which is derived from *Spodoptera frugiperda*.

30 Other suitable vectors for the expression of an olfactory receptor or their peptide fragments or variants in a baculovirus expression system include those described by Chai et al. (1993), Vlasak et al. (1983) and Lenhard et al. (1996).

Mammalian expression vectors will comprise an origin of replication, a suitable promoter and enhancer, and also any necessary ribosome binding sites, polyadenylation site, splice donor and 35 acceptor sites, transcriptional termination sequences, and 5' flanking nontranscribed sequences.

DNA sequences derived from the SV40 viral genome, for example SV40 origin, early promoter,

enhancer, splice and polyadenylation sites may be used to provide the required nontranscribed genetic elements.

Promoters

5 The suitable promoter regions used in the expression vectors according to the present invention are chosen taking into account of the cell host in which the heterologous gene has to be expressed.

10 A suitable promoter may be heterologous with respect to the nucleic acid for which it controls the expression or alternatively can be endogenous to the native polynucleotide containing the coding sequence to be expressed. Additionally, the promoter is generally heterologous with respect to the recombinant vector sequences within which the construct promoter/coding sequence 15 has been inserted.

Thus, the promoter is selected among the group comprising :

15 - an internal or an endogenous promoter, such as the natural promoter associated with the structural gene coding for the desired olfactory receptor polypeptide or the fragment or variant thereof; such a promoter may be completed by a regulatory element derived from the vertebrate host, in particular an activator element;

- a promoter derived from a cytoskeletal protein gene such as the desmin promoter (Bolmont et al., 1990; Zhenlin et al., 1989) or a promoter derived from a gene specifically expressed in epithelial cells and most preferably in olfactory epithelial cells.

20 Preferred bacterial promoters are the LacI, LacZ, the T3 or T7 bacteriophage RNA polymerase promoters, the polyhedrin promoter, or the p10 protein promoter from baculovirus (Kit Novagen) (Smith et al., 1983.; O'Reilly et al., 1992), the lambda P_R promoter or also the trc promoter.

25 Promoter regions can be selected from any desired gene using, for example, CAT (chloramphenicol transferase) vectors and more preferably pKK232-8 and pCM7 vectors. Particularly preferred bacterial promoters include lacI, lacZ, T3, T7, gpt, lambda PR, PL and trp. Eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-L. Selection of a convenient vector and promoter is well within the level of ordinary skill in the art.

30 The choice of a determined promoter, among the above-described promoters is well in the ability of one skill in the art, guided by his knowledge in the genetic engineering technical field, and by being also guided by the book of Sambrook et al. in 1989 or also by the procedures described by Fuller et al. in 1996 (Fuller S.A. et al., 1996).

35 A preferred constitutive promoter that is used is one of the internal promoters that are active in the resting fibroblasts such the promoter of the phosphoglycerate kinase gene (PGK-1). The PGK-1 promoter is either the mouse promoter or the human promoter such as described by Adra et al.(

1987). Other constitutive promoters may also be used such that the beta-actin promoter (Kort et al., 1983) or the vimentin promoter (Rettlez and Basenga, 1987).

The vector containing the appropriate DNA sequence as described above, more preferably a OLF1 to OLF10 coding polynucleotide, can be utilized to transform an appropriate host to allow the 5 expression of the desired polypeptide or polynucleotide.

Other types of vectors

The *in vivo* expression of an olfactory receptor polypeptide encompassed by the invention or a fragment or a variant thereof may be useful in order to correct a genetic defect related to the expression of the native gene in a host organism or to the production of biologically active olfactory 10 receptor proteins.

Consequently, the present invention also deals with recombinant expression vectors mainly designed for the *in vivo* production of a therapeutic peptide fragment by the introduction of the genetic information in the organism of the patient to be treated. This genetic information may be introduced *in vitro* in a cell that has been previously extracted from the organism, the modified cell 15 being subsequently reintroduced in the said organism, directly *in vivo* into the appropriate tissue, and preferably in the olfactory epithelium.

One specific embodiment for a method for delivering the corresponding protein or peptide to the interior of a cell of a vertebrate *in vivo* comprises the step of introducing a preparation comprising a physiologically acceptable carrier and a naked polynucleotide operatively coding for 20 the polypeptide into the interstitial space of a tissue comprising the cell, whereby the naked polynucleotide is taken up into the interior of the cell and has a physiological effect.

In a specific embodiment, the invention provides a composition for the *in vivo* production of an olfactory receptor polypeptide described therein containing a naked polynucleotide operatively coding for an olfactory receptor selected from the group of OLF1 to OLF10 or a fragment or a 25 variant thereof, in solution in a physiologically acceptable carrier and suitable for introduction into a tissue to cause cells of the tissue to express the said protein or polypeptide.

Advantageously, the composition described above is administered locally, near the site in which the expression of the olfactory receptor polypeptide under consideration or a fragment or a variant thereof is sought.

30 The polynucleotide operatively coding for an olfactory receptor polypeptide or a fragment or variant thereof may be a vector comprising the genomic DNA or the complementary DNA (cDNA) coding for the corresponding protein and a promoter sequence allowing the expression of the genomic DNA or the complementary DNA in the desired eukaryotic cells, such as vertebrate cells, specifically mammalian cells.

35 This vector may also contain one origin of replication that allows it to replicate in the eukaryotic host cell such as an origin of replication from a bovine papillomavirus. Alternatively, the vector can contain several, for example two, origins of replication of different origins in order to

allow said vector to replicate in different host cells, typically both in a prokaryotic cell such as *E. coli* and in an eukaryotic cell such as a mammalian epithelial cell, preferably a mammalian olfactory epithelial cell.

Compositions comprising a polynucleotide are described in the PCT application N° WO 5 90/11092 (Vical Inc.) and also in the PCT application N° WO 95/11307 (Institut Pasteur, INSERM, Université d'Ottawa) as well as in the articles of Tacson et al. (1996) and of Huygen et al. (1996).

In another embodiment, the DNA to be introduced is complexed with DEAE-dextran (Pagano et al., 1967) or with nuclear proteins (Kaneda et al., 1989), with lipids (Felgner et al., 1987) or encapsulated within liposomes (Fraley et al., 1980).

10 In another embodiment, the polynucleotide encoding an olfactory receptor polypeptide of the invention or a fragment or a variant thereof may be included in a transfection system comprising polypeptides that promote its penetration within the host cells as it is described in the PCT application WO 95/10534 (Seikagaku Corporation). They can also be encapsulated in polymer microparticles as it is described in the PCT Application No WO 94/27238.

15 The vector according to the present invention may advantageously be administered in the form of a gel that facilitates their transfection into the cells. Such a gel composition may be a complex of poly-L-lysine and lactose, as described by Midoux (1993) or also poloxamer 407 as described by Pastore (1994). Said vector may also be suspended in a buffer solution or be associated with liposomes.

20 The amount of the vector to be injected to the desired host organism vary according to the site of injection. As an indicative dose, it will be injected between 0,1 and 100 µg of the vector in an animal body, preferably a mammal body, for example a mouse body.

25 In another embodiment of the vector according to the invention, said vector may be introduced in vitro in a host cell, preferably in a host cell previously harvested from the animal to be treated and more preferably a somatic cell such as a muscle cell. In a subsequent step, the cell that has been transformed with the vector coding for the desired olfactory receptor polypeptide or the desired fragment or variant thereof is implanted back into the animal body in order to deliver the recombinant protein within the body either locally or systemically.

30 Suitable vectors for the *in vivo* expression of an olfactory receptor polypeptide of the invention or a fragment or a variant thereof are described hereunder.

35 In one specific embodiment, the vector is derived from an adenovirus. Preferred adenoviruses vectors according to the invention are those described by Feldman and Steg (1996) or Ohno et al. (1994). Another preferred recombinant adenovirus according to this specific embodiment of the present invention is the adenovirus described by Ohwada et al. (1996) or the human adenovirus type 2 or 5 (Ad 2 or Ad 5) or an adenovirus of animal origin (French patent application N° FR-93.05954).

Among the adenoviruses of animal origin it can be cited the adenoviruses of canine (CAV2, strain Manhattan or A26/61[ATCC VR-800]), bovine, murine (Mav1, Beard et al., 1980) or simian (SAV).

Preferably, the inventors are using recombinant defective adenoviruses that may be prepared 5 following a technique well-known by one skill in the art, for example as described by Levrero et al. (1991) or by Graham (1984) or in the European patent application N° EP-185.573. Another defective recombinant adenovirus that may be used according to the present invention, as well as a composition of matter containing such a defective recombinant adenovirus, is described in the PCT application N° WO 95/14785.

10 Retrovirus vectors and adeno-associated virus vectors are generally understood to be the recombinant gene delivery system of choice for the transfer of exogenous polynucleotides *in vivo*, particularly to mammals, including humans. These vectors provide efficient delivery of genes into cells, and the transferred nucleic acids are stably integrated into the chromosomal DNA of the host.

15 The use of recombinant retrovirus vectors containing a nucleic acid according to the invention is also encompassed within the scope of the invention. A major prerequisite for the use of retroviruses is to ensure the safety of their use, particularly with regard to the possibility of the spread of wild-type virus in the cell population. The development of specialized cell lines (termed “packaging cells”) which produce only replication defective retroviruses has increased the utility of retroviruses for *in vivo* gene delivery, and defective retroviruses are well characterized for use in 20 gene transfer. Thus, recombinant retroviruses can be constructed in which a part of the retroviral coding sequence (*gag*, *pol*, *env*) has been replaced by nucleic acid encoding an olfactory receptor rendering the retrovirus defective. Protocols for producing recombinant retroviruses and for infecting cells *in vitro* and *in vivo* with such viruses can be found in “Current Protocols in Molecular Biology” (1989).

25 Furthermore, it has been shown that it is possible to limit the infection spectrum of retroviruses and consequently of retroviral-based vectors, by modifying the viral packaging proteins on the surface of the viral particle, as described for example in the PCT Application No WO 93/25234 or in the PCT Application No WO 94/ 06920. For instance, strategies for the modification of the infection spectrum of retroviral vectors include : coupling antibodies specific for cell surface 30 antigens to the viral *env* protein (Julan et al., 1992) or coupling cell surface receptor ligands to the viral *env* protein (Neda et al., 1991). Coupling can be in the form of the chemical cross-linking with a protein or other variety (e.g. lactose to convert the *env* protein to an asialoglycoprotein), as well by generating fusion proteins (e.g. single-chain antibody/*env* fusion proteins). This technique, while useful to limit or otherwise direct the infection to certain tissue types, can also be used to convert an 35 ecotropic vector into an amphotropic vector.

Particularly preferred retroviruses for the preparation or construction of retroviral *in vitro* or *in vitro* gene delivery vehicles of the present invention include retroviruses selected from the group

consisting of Mink-Cell Focus Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis virus and Rous Sarcoma virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley et al., 1976), Abelson (ATCC No VR-999), Friend (ATCC No VR-245), Gross (ATCC No VR-590), Rauscher (ATCC No VR-998) and Moloney Murine Leukemia Virus (ATCC No VR-190; 5 PCT Application No WO 94/24298). Particularly preferred Rous Sarcoma Viruses include Bryan high titer (ATCC Nos VR-334, VR-657, VR-726, VR-659 and VR-728). Another preferred retroviral vector is that described by Roth et al. (Roth J.A. et al., 1996).

Yet another viral vector system that is contemplated by the invention consists in the adeno-associated virus (AAV). Adeno-associated virus is a naturally occurring defective virus that requires 10 another virus, such as an adenovirus or a herpes virus, as a helper virus for efficient replication and a productive life cycle (Muzyczka et al., 1992). It is also one of the few viruses that may integrate its DNA into non-dividing cells, and exhibits a high frequency of stable integration (Flotte et al., 1992; Samulski et al., 1989; McLaughlin et al., 1989). One advantageous feature of AAV derives from its reduced efficacy for transducing primary cells relative to transformed cells.

15 Cell hosts

Another object of the invention consists in host cell that have been transformed or transfected with one of the polynucleotides described therein, and more precisely a polynucleotide comprising the coding sequence of any of the olfactory receptor polypeptide having the amino acid sequence of SEQ ID Nos 12-21 or fragments or variants thereof. Are included host cells that are 20 transformed (prokaryotic cells) or that are transfected (eukaryotic cells) with a recombinant vector such as one of those described above.

A recombinant host cell of the invention comprises any one of the polynucleotides or the recombinant vectors described therein. More particularly, the cell hosts of the present invention can comprise any of the polynucleotides described in the "Coding regions of the olfactory receptor gene" 25 section, "Genomic sequence of olfactory receptor gene" section, the "Oligonucleotide Probes And Primers" section, the "Polynucleotide constructs" section and the "Expression of an OLF1 to OLF10 coding polypeptide" section.

Suitable prokaryotic hosts for transformation include *E. coli*, *Bacillus subtilis*, as well as various species within the genera of *Streptomyces* or *Mycobacterium*. Suitable eukaryotic hosts 30 comprise yeast, insect cells, such as *Drosophila* and Sf9. Various mammalian cell hosts can also be employed to express recombinant protein. Examples of mammalian cell hosts include the COS-7 lines of monkey kidney fibroblasts (Guzman, 1981), and other cell lines capable of expressing a compatible vector, for example the C127, 3T3, CHO, HeLa and BHK cell lines. The selection of an host is within the scope of the one skilled in the art.

35 Preferred cell hosts used as recipients for the expression vectors of the invention are the followings :

a) Prokaryotic host cells : *Escherichia coli* strains (I.E. DH5- α strain) or *Bacillus subtilis*.

b) Eukaryotic host cells : HeLa cells (ATCC N°CCL2; N°CCL2.1; N°CCL2.2). Cv 1 cells (ATCC N°CCL70), COS cells (ATCC N°CRL1650; N°CRL1651), Sf-9 cells (ATCC N°CRL1711).

The constructs in the host cells can be used in a conventional manner to produce the gene product encoded by the recombinant sequence.

5 Following transformation of a suitable host and growth of the host to an appropriate cell density, the selected promoter is induced by appropriate means, such as temperature shift or chemical induction, and cells are cultivated for an additional period.

Cells are typically harvested by centrifugation, disrupted by physical or chemical means, and the resulting crude extract retained for further purification.

10 Microbial cells employed in expression of proteins can be disrupted by any convenient method, including freeze-thaw cycling, sonication, mechanical disruption, or use of cell lysing agents. Such methods are well known by the skill artisan.

Transgenic animals

The terms "transgenic animals" or "host animals" are used herein designate animals that 15 have their genome genetically and artificially manipulated so as to include one of the nucleic acids according to the invention. Preferred animals are non-human mammals and include those belonging to a genus selected from *Mus* (e.g. mice), *Rattus* (e.g. rats) and *Oryctogalrus* (e.g. rabbits) which have their genome artificially and genetically altered by the insertion of a nucleic acid according to the invention.

20 The transgenic animals of the invention all include within a plurality of their cells a cloned recombinant or synthetic DNA sequence, more specifically one of the purified or isolated nucleic acids comprising an olfactory receptor coding sequence selected from the group OLF1 to OLF10 an olfactory receptor regulatory polynucleotide or a DNA sequence encoding an antisense polynucleotide such as described in the present specification.

25 More particularly, transgenic animals according to the invention contain in their somatic cells and/or in their germ line cells any of the polynucleotides described in the "Coding regions of the olfactory receptor gene" section, "Genomic sequence of olfactory receptor gene" section, the "Oligonucleotide Probes And Primers" section, the "Polynucleotide constructs" section and the "Expression of an OLF1 to OLF10 coding polypeptide" section.

30 The replacement of the native genomic olfactory receptor sequence by a defective copy of said sequence may be performed by techniques of gene targeting. Such techniques are notably described by Burright et al. (1997), Bates et al. (1997), Mangiarini et al. (1997), Davies et al. (1997).

Second preferred transgenic animals of the invention have the murine olfactory receptor gene replaced either by a defective copy of the murine olfactory receptor gene or by an interrupted 35 copy of the human olfactory receptor gene. A "defective copy" of a murine or a human olfactory receptor gene, is intended to designate a modified copy of these genes that is not or poorly transcribed in the resulting recombinant host animal or a modified copy of these genes leading to the

absence of synthesis of the corresponding translation product or alternatively leading to a modified and/or truncated translation product lacking the biological activity of the wild type olfactory receptor protein. The altered translation product thus contains amino acid modifications, deletions and substitutions. Modifications and deletions may render the naturally occurring gene nonfunctional, 5 thus leading to a "knockout animal". These transgenic animals are critical for the creation of animal models of human diseases, and for eventual treatment of disorders related to alteration of the olfactory perception of odorant substances or molecules. Examples of such knockout mice are described in the PCT Applications Nos WO 97/34641, WO 96/12792 and WO 98/02354.

The endogenous murine olfactory receptor gene can be interrupted by the insertion, between 10 two contiguous nucleotide of said gene, of a part of all of a marker gene placed under the control of the appropriate promoter, for example the endogenous promoter of the endogenous murine olfactory receptor gene. The marker gene may be the neomycin resistance gene (*neo*) that may be operably linked to the phosphoglycerate kinase-1 (PGK-1) promoter, as described in the PCT Application No WO 98/02534.

15 Thus, the invention is also directed to a transgenic animal contain in their somatic cells and/or in their germ line cells a polynucleotide selected from the following group of polynucleotides:

a) a defective copy of the human olfactory receptor gene;
b) a defective copy of the endogenous olfactory receptor gene, wherein the expression 20 "endogenous olfactory receptor gene" designates an olfactory receptor gene that is naturally present within the genome of the animal host to be genetically modified.

The invention also concerns a method for obtaining transgenic animals, wherein said methods comprise the steps of :

a) replacing the endogenous copy of the animal olfactory receptor gene by a nucleic acid 25 selected from the group consisting of a defective copy of the human olfactory receptor gene and a defective copy of the endogenous olfactory receptor gene in animal cells, preferably embryonic stem cells (ES);
b) introducing the recombinant animal cells obtained at step a) in embryos, notably blastocysts of the animal;
c) selecting the resulting transgenic animals, for example by detecting the defective copy of 30 an olfactory receptor gene with one or several primers or probes according to the invention.

Optionally, the transgenic animals may be bred together in order to obtain homozygous transgenic animals for the defective copy of the olfactory receptor gene introduced.

The transgenic animals of the invention thus contain specific sequences of exogenous 35 genetic material such as the nucleotide sequences described above in detail.

In a preferred embodiment, these transgenic animals may be good experimental models in order to study the diverse pathologies related to disorders associated to alteration of the olfactory

perception of odorant substances or molecules, in particular concerning the transgenic animals within the genome of which has been inserted one or several copies of a polynucleotide encoding a native olfactory receptor protein, or alternatively a mutant olfactory receptor protein.

Third preferred transgenic animals according to the invention contains in their somatic cells 5 and/or in their germ line cells a polynucleotide selected from the following group of polynucleotides :

a) purified or isolated nucleic acid encoding an olfactory receptor polypeptide selected from OLF1 to OLF10, or a polypeptide fragment or variant thereof.

10 b) a purified or isolated nucleic comprising at least 8 consecutive nucleotides of the nucleotide sequence SEQ ID No 1, a nucleotide sequence complementary thereto or a fragment or a variant thereof;

c) a purified or isolated nucleic acid comprising a nucleotide sequence selected from the group of SEQ ID 2-11, a sequence complementary thereto or a fragment or a variant thereof.

15 The transgenic animals of the invention thus contain specific sequences of exogenous genetic material such as the nucleotide sequences described above in detail.

In a first preferred embodiment, these transgenic animals may be good experimental models in order to screen the candidate substance of interest interacting with the olfactory receptor under consideration.

Since it is possible to produce transgenic animals of the invention using a variety of different 20 sequences, a general description will be given of the production of transgenic animals by referring generally to exogenous genetic material. This general description can be adapted by those skilled in the art in order to incorporate the DNA sequences into animals. For more details regarding the production of transgenic animals, and specifically transgenic mice, it may be referred to Sandou et al. (1994) and also to US Patents Nos 4,873,191, issued Oct. 10, 1989, 5,968,766, issued Dec. 16, 25 1997 and 5,387,742, issued Feb. 28, 1995.

Transgenic animals of the present invention are produced by the application of procedures which result in an animal with a genome that incorporates exogenous genetic material which is integrated into the genome. The procedure involves obtaining the genetic material, or a portion thereof, which encodes either a coding sequence, a non-coding polynucleotide or a DNA sequence 30 encoding an antisense polynucleotide of an olfactory receptor selected from the group OLF1 to OLF10 such as described in the present specification.

A recombinant polynucleotide of the invention is inserted into an embryonic or ES stem cell line. The insertion is made using electroporation. The cells subjected to electroporation are screened (e.g. Southern blot analysis) to find positive cells which have integrated the exogenous recombinant 35 polynucleotide into their genome. An illustrative positive-negative selection procedure that may be used according to the invention is described by Mansour et al. (1988). Then, the positive cells are isolated, cloned and injected into 3.5 days old blastocysts from mice. The blastocysts are then

inserted into a female host animal and allowed to grow to term. The offsprings of the female host are tested to determine which animals are transgenic e.g. include the inserted exogenous DNA sequence and which are wild-type.

Thus, the present invention also concerns a transgenic animal containing a nucleic acid, a recombinant expression vector or a recombinant host cell according to the invention.

Recombinant Cell Lines Derived From The Transgenic Animals Of The Invention.

A further object of the invention comprises recombinant host cells obtained from a transgenic animal described herein. In one embodiment the invention encompasses cells derived from non-human host mammals and animals comprising a recombinant vector of the invention or an olfactory receptor gene disrupted by homologous recombination with a knock out vector.

Recombinant cell lines may be established *in vitro* from cells obtained from any tissue of a transgenic animal according to the invention, for example by transfection of primary cell cultures with vectors expressing *onc*-genes such as SV40 large T antigen, as described by Chou (1989) and Shay et al.(1991).

15 F. METHODS FOR SCREENING SUBSTANCES OR MOLECULES INTERACTING WITH AN OLFACTORY RECEPTOR PROTEIN

The present invention pertains to methods for screening substances of interest, in particular odorant substances or molecules that interact with an olfactory receptor protein selected from the group consisting of OLF1 to OLF10, or one peptide fragment or variant thereof. In one embodiment, the candidate substance is devoid of odorant propriety but it is able to bind the olfactory receptor and to trigger the transduction of signals.

For the purpose of the present invention, a ligand means a molecule, such as a protein, a peptide, an antibody or any synthetic chemical compound capable of binding to the olfactory receptor protein or one of its fragments or variants or to modulate the expression of the polynucleotide coding for olfactory receptor or a fragment or variant thereof.

In the ligand screening method according to the present invention, a biological sample or a defined molecule to be tested as a putative ligand of the olfactory receptor protein is brought into contact with the corresponding purified olfactory receptor protein, for example the corresponding purified recombinant olfactory receptor protein produced by a recombinant cell host as described herein, in order to form a complex between this protein and the putative ligand molecule to be tested.

As an illustrative example, to study the interaction of the olfactory receptor protein, or a fragment comprising comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments" with drugs or small molecules, such as molecules generated through combinatorial chemistry approaches, the microdialysis coupled to HPLC method described by Wang et al. (1997) or the affinity capillary electrophoresis method described by Bush et al. (1997) can be used.

In further methods, peptides, drugs, fatty acids, lipoproteins, or small molecules which interact with the olfactory receptor protein, or a fragment comprising any of the fragments described in the section "OLF1 to OLF10 proteins and polypeptide fragments" may be identified using assays such as the following. The molecule to be tested for binding is labeled with a detectable label, such 5 as a fluorescent, radioactive, or enzymatic tag and placed in contact with immobilized olfactory receptor protein, or a fragment thereof under conditions which permit specific binding to occur, such as affinity columns. In some embodiments, chimeric proteins containing the olfactory receptor protein fused to proteins facilitating purification, such as glutathione S transferase (GST) are used. After removal of non-specifically bound molecules, bound molecules are detected using appropriate 10 means.

In one embodiment, proteins, peptides, carbohydrates, lipids, or small molecules generated by combinatorial chemistry interacting with the olfactory receptor protein, or a fragment or a variant thereof can also be screened by using an Optical Biosensor as described in Edwards and Leatherbarrow (1997) and also in Szabo et al. (1995). The main advantage of the method is that it 15 allows the determination of the association rate between the olfactory receptor protein and molecules interacting with the olfactory receptor protein. It is thus possible to select specifically ligand molecules interacting with the olfactory receptor protein, or a fragment thereof, through strong or conversely weak association constants.

Another object of the present invention comprises methods and kits for the screening of 20 candidate substances that interact with olfactory receptor polypeptide.

The present invention pertains to methods for screening substances of interest that interact with an olfactory receptor protein or one fragment or variant thereof. By their capacity to bind covalently or non-covalently to an olfactory receptor protein or to a fragment or variant thereof, these substances or molecules may be advantageously used both *in vitro* and *in vivo*. *In vitro*, said 25 interacting molecules may be used as detection means in order to identify the presence of an olfactory receptor protein in a sample, preferably a biological sample.

A first method for the screening of a candidate substance interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, or fragments or variants thereof, comprises the following steps :

- 30 a) providing a polypeptide selected from the group consisting of the polypeptides comprising, consisting essentially of, or consisting of the amino acid sequences of SEQ ID Nos 12-21, or a peptide fragment or a variant thereof;
- b) obtaining a candidate substance;
- c) bringing into contact said polypeptide with said candidate substance; and
- 35 d) detecting the complexes formed between said polypeptide and said candidate substance.

Various candidate substances or molecules can be assayed for interaction with an olfactory receptor polypeptide. These substances or molecules include, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as polypeptides. When the candidate substance or molecule comprises a polypeptide, this polypeptide may be the resulting 5 expression product of either a phage clone belonging to a phage-based random peptide library, or of a cDNA library cloned in a vector suitable for performing a two-hybrid screening assay.

In one embodiment of the screening method defined above, the complexes formed between the polypeptide and the candidate substance are further incubated in the presence of a polyclonal or a monoclonal antibody that specifically binds to the olfactory receptor protein of the invention under 10 consideration or to said peptide fragment or variant thereof.

In another embodiment of the present screening method, increasing concentrations of a substance competing for binding to the olfactory receptor with the considered candidate substance is added, simultaneously or prior to the addition of the candidate substance or molecule, when performing step c) of said method. By this technique, the detection and optionally the quantification 15 of the complexes formed between the olfactory receptor protein or the peptide fragment or variant thereof and the candidate substance or molecule to be screened allows the one skilled in the art to determine the affinity value of said substance or molecule for said olfactory receptor protein or the peptide fragment or variant thereof.

The olfactory receptor selected from the group consisting of OLF1 to OLF10, or a peptide 20 fragment or a variant thereof, can be overexpressed and purified in a bacterial system such as *E. coli* as described in Kiefer et al. (1996) and Tucker et al. (1996). The olfactory receptor coding sequence can be fused to its N-terminus with GST (Glutathione S transferase) or MBP (Maltose Binding Protein) and to its C-terminus with poly-histidine tag, Bio tag or Strep tag for facilitating the purification of the expressed protein. The Bio tag is 13 amino acid residues long, is biotinylated *in* 25 *vivo* in *E. coli*, and will therefore bind to both avidin and streptavidin. The Strep tag is 9 amino acid residues long and binds specifically to streptavidin, but not to avidin. Therewith, a purification step by affinity can be carried out based on the interaction of a poly-histidine tail with immobilized metal ions, of the biotinylated Bio tag with monomeric avidin, of the Strep tag with streptavidin, of the GST segment with the glutathione, or of the MBP segment with the maltose. Thioredoxin can be 30 eventually inserted between the receptor C-terminus and the tag and could increase the expression level. The fusion protein is solubilized in 1% N-laurroyl sarcosine, and 0.2 % digitonin is added. It is purified by affinity chromatography. The MBP, GST or tag segment can be then removed. After the olfactory receptor protein purification, sarcosyl can be replaced with digitonin which is a detergent widely used to stabilize the G protein-coupled receptors. The purified receptor is reconstituted into 35 lipid vesicles preferably composed of phosphatidylcholine: phosphatidylglycerol (4:1) by adding the lipid dissolved in dodecyl maltoside and removing the detergent.

The olfactory receptor selected from the group consisting of OLF1 to OLF10, or a peptide fragment or a variant thereof, can also be overexpressed and purified in a baculovirus/Sf9 system as described in Nekrasova et al. (1996). The olfactory receptor gene, or a fragment thereof, is preferably expressed with a "flag" peptide epitope tag and/or a poly-histidine tag to either its N- or 5 C-terminus for facilitating the purification of the expressed protein. Therefore, the olfactory receptor gene, or a fragment or a variant thereof, is preferably subcloned into the baculovirus transfer vector pAcSGHisNT to create constructs that encoded olfactory receptor with amino-terminal poly-histidine tag. The resulting transfer vector is transfected preferably with BaculoGold DNA into Sf9 cells. The expressed olfactory receptors are then solubilized either in 1 % N-lauryl sarcosine or 1.5 10 % lysophosphatidylcholine, but preferably in lysophosphatidylcholine. After solubilization, the olfactory receptors are purified by affinity chromatography on nickel nitrilotriacetic acid resin and by cation-exchange chromatography with carboxymethyl sepharose cation-exchange column. The tag segment can be then removed. The purified receptor is reconstituted into lipid vesicles preferably composed of dimyristoylglycerophosphocholine, cholesterol, dialmitoylglycerophosphoserine and 15 dipalmitoylglycerophosphoethanolamine (in molecular ratio 54:35:10:1)

Once the olfactory receptor protein or one of its peptide fragments or variants has been obtained as described above, candidate substances or molecules can then be assayed for their capacity to bind thereto.

The candidate substance or molecule to be assayed for interacting with an olfactory receptor 20 of the invention may be of diverse nature, including, without being limited to, natural or synthetic organic compounds or molecules of biological origin such as peptide. It can comprise aromatic or aliphatic compounds with various functional groups such as alcohol, aldehyde, ester, ether, ketone, carboxylic, amine. An example of a substance panel which can be used is provided by Zhao et al. (1998).

25 The screening of substances or molecules interacting with an olfactory receptor, or a fragment thereof, is carried out by photoaffinity labeling experiments described in Kiefer et al. (1996). The odorant is labeled, preferably radiolabeled, and incubated with lipid vesicles including the purified olfactory receptor. The odorants bound to the olfactory receptors are crosslinked by exposure to ultraviolet light. Then, the samples are subjected to SDS polyacrylamide gel 30 electrophoresis. Proteins are visualized by Coomassie-blue staining and the odorants are revealed, preferably by autoradiography. In another embodiment, the proteins can be visualized by Western Blot with a polyclonal or monoclonal antibody that specifically binds to the olfactory receptor under consideration. Once a substance binding to the considered olfactory receptor is identified, the binding specificity of this substance is confirmed with competition experiments demonstrating that 35 increasing concentrations of unlabeled ligand accomplish a dose-dependent displacement of the radioactive ligand.

The identification of a first substance specific to one of the olfactory receptors of the present invention facilitates the screening of other substances. Indeed, the binding capacity of the screened substances to this olfactory receptor can be carried out through a competition experiments against the first identified substance which is labeled.

5 The invention also pertains to kits useful for performing the hereinbefore described screening method. Preferably, such kits comprise a polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21 or a peptide fragment or a variant thereof, and optionally means useful to detect the complex formed between the considered olfactory receptor polypeptide or its peptide fragment or variant and the candidate substance. In a 10 preferred embodiment, the kit can comprise an already identified substance specific of the olfactory receptor under consideration which is labeled, preferably radiolabeled, and a monoclonal or polyclonal antibody directed against the considered olfactory receptor.

A second screening method embodiment consists of a method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of 15 SEQ ID Nos 12-21, wherein said method comprises :

- a) providing a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group comprising, consisting essentially of, or consisting the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, or variants or fragments thereof;
- 20 b) preparing membrane extracts of said recombinant eukaryotic host cell;
- c) bringing into contact the membrane extracts prepared at step b) with a selected ligand molecule; and
- d) detecting the production level of second messengers metabolites.

The baculovirus-Sf9 cell system enables a foreign DNA encoding an olfactory receptor 25 selected from the group consisting of OLF1 to OLF10, or a peptide fragment or a variant thereof, to be expressed with high efficiency. Moreover, it can be used to couple a heterologous expressed olfactory receptor to the second messenger cascades. Therefore, the binding specificity of an olfactory receptor can be assessed through an assay of odorant-induced generation of cAMP or inositol triphosphate (InsP3) described in Raming et al. (1993).

30 Briefly, a cell line derived from Sf9 is infected by baculovirus, such as baculovirus transfer vector pVL1393, harboring DNA encoding the olfactory receptor or a fragment thereof downstream from a strong promoter, preferably the polyhedrin promoter. Recombinant virus are purified and used to infect 1.5×10^8 Sf9 cells in 100 ml spinner cultures at high multiplicity of infection. Cells are collected after a postinfection delay, preferably 48 h, and membrane fractions are isolated as follow.

35 Cells are pelleted (at 250g for 10 min at 4°C), washed with Ringer solution (120 mM NaCl, 5 mM KCl, 1.6 mM K₂HPO₄, 1.2 mM MgSO₄, 25 mM NaHCO₃, 5 mM glucose, pH7.4) and disrupted using a glass homogenizer in homogenization buffer (10 mM Tris-HCl, pH 8.0, 2 mM

EGTA, 3 mM MgCl₂) containing antiproteases. The homogenate is centrifuged and the pellet is washed. Supernatants are centrifuged at 33,000g for 20 min. The final pellet is resuspended in homogenization buffer and the protein concentration is determined.

Assay of odorant substance-induced generation of second messengers cAMP and InsP3 is 5 performed as follow. Suspensions of Sf9 cell membrane preparations (300 µg protein) are rapidly mixed with a stimulation buffer (200 mM NaCl, 10 mM EGTA, 50 mM MOPS, 2.5 mM MgCl₂, 1 mM DTT, 0.05 % Na-cholate, 1 mM ATP, 1 µM GTP, and 0.02 µM free Ca²⁺) containing the candidate substances at the appropriate concentrations. The reaction is stopped after a short time, preferably 1 sec, by injecting 10 % Perchloric acid. Quenched samples are assayed for second 10 messenger concentrations. The quenched and cooled samples are vortexed and centrifuged for 5 min at 2500g at 4°C. 400 µl of the supernatants are transferred to a separate tube containing 100 µl of 10 mM EDTA (pH 7). The sample are neutralized by adding 500 µl of a 1:1 (v/v) mixture of 1,1,2 trichlorofluoroethane, followed by thorough mixing. After centrifugation for 2 min at 500g, three 15 phases are obtained. The upper phase, which contains all water soluble components, is used for carrying out the concentration measurements. cAMP and InsP3 concentrations are determined according to the procedure of Steiner et al. (1972) and Palmer et al. (1989), respectively.

The invention also concerns a kit for the screening of odorant ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, wherein said kit comprises :

20 a) a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group comprising, consisting essentially of, or consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21 or variants or fragments thereof; and
25 b) optionally, reagents necessary for the measurement of second messenger metabolites in a sample.

The screening of substances or molecules interacting with an olfactory receptor, or a fragment thereof, can also be carried out through the measurement of the increase of the response to odorants in an olfactory epithelium overexpressing an olfactory receptor selected from the group consisting of OLF1 to OLF10, or a peptide fragment or a variant thereof, as described in Zhao et al. 30 (1998). The response is assessed by electro-olfactogram which measures a transepithelial potential resulting from the summed activity of many olfactory neurons. In order to overexpress the olfactory receptor, or a fragment thereof, in an olfactory epithelium, an adenovirus containing the olfactory receptor gene is generated. To aid in electro-olfactogram electrode placements, the olfactory receptor coding sequence is preferably combined in the adenovirus with the physiological marker 35 green fluorescent protein (GFP) in such manner that the two proteins are simultaneously expressed. The olfactory epithelium of an animal, preferably of a rat, is infected by the adenovirus. Animals are killed 3 to 8 days after infection and the nasal cavity is opened, exposing the medial surface of the

nasal turbinates. Under fluorescent illumination, the GFP clearly marked the pattern of viral infection and olfactory receptor expression. Odorant substance are applied to the olfactory epithelium in the vapor phase by injecting a pressurized pulse of odorant vapor into a continuous stream of humidified clean air. Electro-olfactogram recordings are obtained with a glass capillary 5 electrode placed on the surface of the epithelium and connected to a differential amplifier. The olfactory receptor specificity is assessed from the increase of response in infected animals compared to uninfected animals. To account for the variability between animals, a standard odorant to which all other odorant responses are normalized is used.

A third screening method embodiment consists of a method for the screening of ligand 10 molecules interacting with an olfactory receptor polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, wherein said method comprises :

- 15 a) providing an adenovirus containing a nucleic acid encoding a polypeptide selected from the group comprising, consisting essentially of, or consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21, or variants or fragments thereof;
- b) infecting an olfactory epithelium with said adenovirus;
- c) bringing into contact the olfactory epithelium b) with a selected ligand molecule; and
- d) detecting the increase of the response to said ligand molecule.

20 G. METHODS FOR INHIBITING THE EXPRESSION OF AN OLFACTORY RECEPTOR GENE

Other therapeutic compositions according to the present invention comprise advantageously an oligonucleotide fragment of the nucleic sequence of olfactory receptor as an antisense tool or a triple helix tool that inhibits the expression of the corresponding olfactory receptor gene. A 25 preferred fragment of the nucleic sequence of olfactory receptor comprises an allele of at least one of the biallelic markers A1 to A13.

Antisense Approach

Preferred methods using antisense polynucleotide according to the present invention are the procedures described by Sczakiel et al.(1995).

30 Preferred antisense polynucleotides are described in the section entitled "Nuclear Antisense DNA Constructs".

The antisense nucleic acids should have a length and melting temperature sufficient to permit formation of an intracellular duplex having sufficient stability to inhibit the expression of the olfactory receptor mRNA in the duplex. Strategies for designing antisense nucleic acids suitable for 35 use in gene therapy are disclosed in Green et al., (1986) and Izant and Weintraub, (1984).

In some strategies, antisense molecules are obtained by reversing the orientation of the olfactory receptor coding region with respect to a promoter so as to transcribe the opposite strand from that which is normally transcribed in the cell. The antisense molecules may be transcribed using in vitro transcription systems such as those which employ T7 or SP6 polymerase to generate 5 the transcript. Another approach involves transcription of olfactory receptor antisense nucleic acids in vivo by operably linking DNA containing the antisense sequence to a promoter in a suitable expression vector.

Alternatively, suitable antisense strategies are those described by Rossi et al.(1991), in the International Applications Nos. WO 94/23026, WO 95/04141, WO 92/18522 and in the European 10 Patent Application No. EP 0 572 287 A2

An alternative to the antisense technology that is used according to the present invention comprises using ribozymes that will bind to a target sequence via their complementary polynucleotide tail and that will cleave the corresponding RNA by hydrolyzing its target site (namely "hammerhead ribozymes"). Briefly, the simplified cycle of a hammerhead ribozyme 15 comprises (1) sequence specific binding to the target RNA via complementary antisense sequences; (2) site-specific hydrolysis of the cleavable motif of the target strand; and (3) release of cleavage products, which gives rise to another catalytic cycle. Indeed, the use of long-chain antisense polynucleotide (at least 30 bases long) or ribozymes with long antisense arms are advantageous. A preferred delivery system for antisense ribozyme is achieved by covalently linking these antisense 20 ribozymes to lipophilic groups or to use liposomes as a convenient vector. Preferred antisense ribozymes according to the present invention are prepared as described by Szakiel et al.(1995), the specific preparation procedures being referred to in said article.

Triple Helix Approach

The olfactory receptor genomic DNA may also be used to inhibit the expression of the 25 olfactory receptor gene based on intracellular triple helix formation.

Triple helix oligonucleotides are used to inhibit transcription from a genome. They are particularly useful for studying alterations in cell activity when it is associated with a particular gene.

Similarly, a portion of the olfactory receptor genomic DNA can be used to study the effect 30 of inhibiting olfactory receptor transcription within a cell. Traditionally, homopurine sequences were considered the most useful for triple helix strategies. However, homopyrimidine sequences can also inhibit gene expression. Such homopyrimidine oligonucleotides bind to the major groove at homopurine:homopyrimidine sequences. Thus, both types of sequences from the olfactory receptor genomic DNA are contemplated within the scope of this invention.

35 To carry out gene therapy strategies using the triple helix approach, the sequences of the olfactory receptor genomic DNA are first scanned to identify 10-mer to 20-mer homopyrimidine or homopurine stretches which could be used in triple-helix based strategies for inhibiting olfactory

receptor expression. Following identification of candidate homopyrimidine or homopurine stretches, their efficiency in inhibiting olfactory receptor expression is assessed by introducing varying amounts of oligonucleotides containing the candidate sequences into tissue culture cells which express the olfactory receptor gene.

5 The oligonucleotides can be introduced into the cells using a variety of methods known to those skilled in the art, including but not limited to calcium phosphate precipitation, DEAE-Dextran, electroporation, liposome-mediated transfection or native uptake.

10 Treated cells are monitored for altered cell function or reduced olfactory receptor expression using techniques such as Northern blotting, RNase protection assays, or PCR based strategies to monitor the transcription levels of the olfactory receptor gene in cells which have been treated with the oligonucleotide.

15 The oligonucleotides which are effective in inhibiting gene expression in tissue culture cells may then be introduced in vivo using the techniques described above in the antisense approach at a dosage calculated based on the in vitro results, as described in antisense approach.

20 In some embodiments, the natural (beta) anomers of the oligonucleotide units can be replaced with alpha anomers to render the oligonucleotide more resistant to nucleases. Further, an intercalating agent such as ethidium bromide, or the like, can be attached to the 3' end of the alpha oligonucleotide to stabilize the triple helix. For information on the generation of oligonucleotides suitable for triple helix formation see Griffin et al.(1989).

20 H. COMPUTER-RELATED EMBODIMENTS

25 As used herein the term "nucleic acid codes of the invention" encompass the nucleotide sequences comprising, consisting essentially of, or consisting of any of the polynucleotides described in the "Coding Regions of the olfactory receptor gene" section, "Genomic sequence of the olfactory receptor gene" section and the "Oligonucleotide Probes And Primers" section, or variants thereof, or complementary sequences thereto. Homologous sequences refer to a sequence having at least 99%, 98%, 97%, 96%, 95%, 90%, 85%, 80%, or 75% homology to these contiguous spans. Homology may be determined using any method described herein, including BLAST2N with the default parameters or with any modified parameters. Homologous sequences also may include RNA sequences in which uridines replace the thymines in the nucleic acid codes of the invention.

30 As used herein the term "polypeptide codes of the invention" encompass the polypeptide sequences comprising any of the polypeptides described in the " OLF1 to OFL10 proteins and polypeptide fragments".

35 It will be appreciated that the nucleic acid and polypeptide codes of the invention can be represented in the traditional single character format or three letter format respectively (See the inside back cover of Stryer, Lubert. *Biochemistry*, 3rd edition. W. H Freeman & Co., New York.) or in any

other format or code which records the identity of the nucleotides or the amino acid respectively in a sequence.

It will be appreciated by those skilled in the art that the nucleic acid codes of the invention and polypeptide codes of the invention can be stored, recorded, and manipulated on any medium which can 5 be read and accessed by a computer. As used herein, the words "recorded" and "stored" refer to a process for storing information on a computer medium. A skilled artisan can readily adopt any of the presently known methods for recording information on a computer readable medium to generate manufactures comprising one or more of the nucleic acid codes of the invention, or one or more of the 10 polypeptide codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 nucleic acid codes of the invention. Another aspect of the present invention is a computer readable medium having recorded thereon at least 2, 5, 10, 15, 20, 25, 30, or 50 polypeptide codes of the invention.

Computer readable media include magnetically readable media, optically readable media, electronically readable media and magnetic/optical media. For example, the computer readable media 15 may be a hard disc, a floppy disc, a magnetic tape, CD-ROM, DVD, RAM, or ROM as well as other types of other media known to those skilled in the art.

Embodiments of the present invention include systems, particularly computer systems which contain the sequence information described herein. As used herein, "a computer system" refers to the hardware components, software components, and data storage components used to store and/or analyze 20 the nucleotide sequences of the nucleic acid codes of the invention, the amino acid sequences of the polypeptide codes of the invention, or other sequences. The computer system preferably includes the computer readable media described above, and a processor for accessing and manipulating the sequence data.

In some embodiments, the computer system may further comprise a sequence comparer for 25 comparing the nucleic acid codes or polypeptide codes of the invention stored on a computer readable medium to reference nucleotide sequences stored on a computer readable medium. A "sequence comparer" refers to one or more programs which are implemented on the computer system to compare a nucleotide or polypeptide sequence with other nucleotide or polypeptide sequences and/or compounds including but not limited to peptides, peptidomimetics, and chemicals the sequences or structures of 30 which are stored within the data storage means. For example, the sequence comparer may compare the nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention stored on a computer readable medium to reference sequences stored on a computer readable medium to identify homologies, motifs implicated in biological function, or structural motifs. The various sequence comparer programs identified elsewhere in this patent 35 specification are particularly contemplated for use in this aspect of the invention.

Accordingly, one aspect of the present invention is a computer system comprising a processor, a data storage device having stored thereon a nucleic acid code of the invention or a

polypeptide code of the invention, a data storage device having retrievably stored thereon reference nucleotide sequences or polypeptide sequences to be compared to the nucleic acid code of the invention or polypeptide code of the invention and a sequence comparer for conducting the comparison. The sequence comparer may indicate a homology level between the sequences

5 compared or identify structural motifs in the nucleic acid code of the invention and polypeptide codes of the invention or it may identify structural motifs in sequences which are compared to these nucleic acid codes and polypeptide codes. In some embodiments, the data storage device may have stored thereon the sequences of at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or polypeptide codes of the invention.

10 Another aspect of the present invention is a method for determining the level of homology between a nucleic acid code of the invention and a reference nucleotide sequence, comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through the use of a computer program which determines homology levels and determining homology between the nucleic acid code and the reference nucleotide sequence with the computer program. The computer program

15 may be any of a number of computer programs for determining homology levels, including those specifically enumerated herein, including BLAST2N with the default parameters or with any modified parameters. The method may be implemented using the computer systems described above. The method may also be performed by reading 2, 5, 10, 15, 20, 25, 30, or 50 of the above described nucleic acid codes of the invention through the use of the computer program and determining homology

20 between the nucleic acid codes and reference nucleotide sequences.

Alternatively, the computer program may be a computer program which compares the nucleotide sequences of the nucleic acid codes of the present invention, to reference nucleotide sequences in order to determine whether the nucleic acid code of the invention differs from a reference nucleic acid sequence at one or more positions. Optionally such a program records the length and

25 identity of inserted, deleted or substituted nucleotides with respect to the sequence of either the reference polynucleotide or the nucleic acid code of the invention. In one embodiment, the computer program may be a program which determines whether the nucleotide sequences of the nucleic acid codes of the invention contain one or more single nucleotide polymorphisms (SNP) with respect to a reference nucleotide sequence. These single nucleotide polymorphisms may each comprise a single

30 base substitution, insertion, or deletion.

Another aspect of the present invention is a method for determining the level of homology between a polypeptide code of the invention and a reference polypeptide sequence, comprising the steps of reading the polypeptide code of the invention and the reference polypeptide sequence through use of a computer program which determines homology levels and determining homology between the

35 polypeptide code and the reference polypeptide sequence using the computer program.

Accordingly, another aspect of the present invention is a method for determining whether a nucleic acid code of the invention differs at one or more nucleotides from a reference nucleotide

sequence comprising the steps of reading the nucleic acid code and the reference nucleotide sequence through use of a computer program which identifies differences between nucleic acid sequences and identifying differences between the nucleic acid code and the reference nucleotide sequence with the computer program. In some embodiments, the computer program is a program 5 which identifies single nucleotide polymorphisms. The method may be implemented by the computer systems described above. The method may also be performed by reading at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention and the reference nucleotide sequences through the use of the computer program and identifying differences between the nucleic acid codes and the reference nucleotide sequences with the computer program.

10 An "identifier" refers to one or more programs which identifies certain features within the above-described nucleotide sequences of the nucleic acid codes of the invention or the amino acid sequences of the polypeptide codes of the invention.

In one embodiment, the identifier may comprise a molecular modeling program which determines the 3-dimensional structure of the polypeptides codes of the invention. In some 15 embodiments, the molecular modeling program identifies target sequences that are most compatible with profiles representing the structural environments of the residues in known three-dimensional protein structures. (See, e.g., Eisenberg et al., U.S. Patent No. 5,436,850 issued July 25, 1995). In another technique, the known three-dimensional structures of proteins in a given family are superimposed to define the structurally conserved regions in that family. This protein modeling 20 technique also uses the known three-dimensional structure of a homologous protein to approximate the structure of the polypeptide codes of the invention. (See e.g., Srinivasan, et al., U.S. Patent No. 5,557,535 issued September 17, 1996). Conventional homology modeling techniques have been used routinely to build models of proteases and antibodies. (Sowdhamini et al., (1997)).

Comparative approaches can also be used to develop three-dimensional protein models when the 25 protein of interest has poor sequence identity to template proteins. In some cases, proteins fold into similar three-dimensional structures despite having very weak sequence identities. For example, the three-dimensional structures of a number of helical cytokines fold in similar three-dimensional topology in spite of weak sequence homology.

The recent development of threading methods now enables the identification of likely 30 folding patterns in a number of situations where the structural relatedness between target and template(s) is not detectable at the sequence level. Hybrid methods, in which fold recognition is performed using Multiple Sequence Threading (MST), structural equivalencies are deduced from the threading output using a distance geometry program DRAGON to construct a low resolution model, and a full-atom representation is constructed using a molecular modeling package such as 35 QUANTA. According to this 3-step approach, candidate templates are first identified by using the novel fold recognition algorithm MST, which is capable of performing simultaneous threading of multiple aligned sequences onto one or more 3-D structures. In a second step, the structural

equivalencies obtained from the MST output are converted into interresidue distance restraints and fed into the distance geometry program DRAGON, together with auxiliary information obtained from secondary structure predictions. The program combines the restraints in an unbiased manner and rapidly generates a large number of low resolution model confirmations. In a third step, these 5 low resolution model confirmations are converted into full-atom models and subjected to energy minimization using the molecular modeling package QUANTA. (See e.g., Aszódi et al., (1997)).

The results of the molecular modeling analysis may then be used in rational drug design techniques to identify agents which modulate the activity of the polypeptide codes of the invention.

Accordingly, another aspect of the present invention is a method of identifying a feature 10 within the nucleic acid codes of the invention or the polypeptide codes of the invention comprising reading the nucleic acid code(s) or the polypeptide code(s) through the use of a computer program which identifies features therein and identifying features within the nucleic acid code(s) or polypeptide code(s) with the computer program. In one embodiment, computer program comprises a computer program which identifies open reading frames. In a further embodiment, the computer 15 program identifies structural motifs in a polypeptide sequence. In another embodiment, the computer program comprises a molecular modeling program. The method may be performed by reading a single sequence or at least 2, 5, 10, 15, 20, 25, 30, or 50 of the nucleic acid codes of the invention or the polypeptide codes of the invention through the use of the computer program and identifying features within the nucleic acid codes or polypeptide codes with the computer program.

20 The nucleic acid codes of the invention or the polypeptide codes of the invention may be stored and manipulated in a variety of data processor programs in a variety of formats. For example, they may be stored as text in a word processing file, such as MicrosoftWORD or WORDPERFECT or as an ASCII file in a variety of database programs familiar to those of skill in the art, such as DB2, SYBASE, or ORACLE. In addition, many computer programs and databases may be used as sequence 25 comparers, identifiers, or sources of reference nucleotide or polypeptide sequences to be compared to the nucleic acid codes of the invention or the polypeptide codes of the invention. The following list is intended not to limit the invention but to provide guidance to programs and databases which are useful with the nucleic acid codes of the invention or the polypeptide codes of the invention. The programs and databases which may be used include, but are not limited to: MacPattern (EMBL), DiscoveryBase 30 (Molecular Applications Group), GeneMine (Molecular Applications Group), Look (Molecular Applications Group), MacLook (Molecular Applications Group), BLAST and BLAST2 (NCBI), BLASTN and BLASTX (Altschul et al, 1990), FASTA (Pearson and Lipman, 1988), FASTDB (Brutlag et al., 1990), Catalyst (Molecular Simulations Inc.), Catalyst/SHAPE (Molecular Simulations Inc.), Cerius².DBAccess (Molecular Simulations Inc.), HypoGen (Molecular Simulations Inc.), Insight 35 II, (Molecular Simulations Inc.), Discover (Molecular Simulations Inc.), CHARMM (Molecular Simulations Inc.), Felix (Molecular Simulations Inc.), DelPhi, (Molecular Simulations Inc.), QuanteMM, (Molecular Simulations Inc.), Homology (Molecular Simulations Inc.), Modeler

(Molecular Simulations Inc.), ISIS (Molecular Simulations Inc.), Quanta/Protein Design (Molecular Simulations Inc.), WebLab (Molecular Simulations Inc.), WebLab Diversity Explorer (Molecular Simulations Inc.), Gene Explorer (Molecular Simulations Inc.), SeqFold (Molecular Simulations Inc.), the EMBL/Swissprotein database, the MDL Available Chemicals Directory database, the MDL Drug

5 Data Report data base, the Comprehensive Medicinal Chemistry database, Derwents's World Drug Index database, the BioByteMasterFile database, the Genbank database, and the Genseqn database. Many other programs and data bases would be apparent to one of skill in the art given the present disclosure.

Motifs which may be detected using the above programs include sequences encoding
10 leucine zippers, helix-turn-helix motifs, glycosylation sites, ubiquitination sites, alpha helices, and beta sheets, signal sequences encoding signal peptides which direct the secretion of the encoded proteins, sequences implicated in transcription regulation such as homeoboxes, acidic stretches, enzymatic active sites, substrate binding sites, and enzymatic cleavage sites.

15 Throughout this application, various publications, patents and published patent applications are cited. The disclosures of these publications, patents and published patent specification referenced in this application are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

EXAMPLES

20 **EXAMPLE 1: LOCALIZATION OF THE OLFACTORY RECEPTOR GENE OLF3 AND OLF5 ON THE HUMAN CHROMOSOMES.**

Metaphase chromosome preparation

Metaphase chromosomes were prepared from phytohemagglutinin (PHA)-stimulated blood cell donors. PHA stimulated lymphocytes from healthy males were cultured for 72 h in RPMI-1640
25 medium. For synchronization, methotrexate (10 μ M) was added for 17 h, followed by addition of 5-bromodeoxyuridine (5-BrdU, 0.1 mM) for 6 h. Colcemid (1 mg/ml) was added for the last 15 min before harvesting the cells. Cells were collected, washed in RPMI, incubated with a hypotonic solution of KCl (75 mM) at 37°C for 15 min and fixed in three changes of methanol:acid acetic (3:1). The cell suspension was dropped onto a glass slide, air-dried and kept in darkness at -20°C
30 until use.

Probes:

- The BAC H0526H04 containing Olf3 and Olf5 genes was used to generate probe by Alu-PCR. PCR amplification of BAC recombinant DNA (50 ng) was carried out as described by Romana et al. (1993).

- Two DNA fragments carrying respectively Olf3 and Olf5 sequences were generated by long range PCR with specific primers (SEQ ID 96-99) and used as probes to confirm the localization of each genes. Olf3 and Olf5 amplicons are respectively 2.8 kb and 3.2 kb fragments.

Probes were labeled by nick translation with bio-16-dUTP (Boehringer Mannheim), and 5 purified over a Sephadex G50 column.

Fluorescence In Situ Hybridization

To determine the chromosomal localization of both genes, the BAC probe was initially hybridized to human metaphase cells. When biotinylated PCR products of BAC DNA were used in hybridization experiment, 75 ng of probe was precipitated with 75 µg of competitor DNA (human 10 Cot1 DNA, GIBCO-BRL) and resuspended in 10 µl of hybridization buffer (50% formamide, 2 X SSC, 10% dextran sulfate, 1 mg/ml sonicated herring DNA, pH 7). When long range PCR products of Olf3 or Olf5 genes were used as probe, 5 ng of biotinylated probe were mixed with 5 µg of human Cot1 DNA. Prior to hybridization, the probe was denatured at 70°C for 10 min and preannealed at 37°C for 2 h.

15 Slides were treated for 1 h at 37°C with Rnase A (100 µg/ml), rinsed three times in 2 X SSC and dehydrated in an ethanol serie. Chromosome preparations were denatured in 70% formamide, 2 X SSC (pH 7), for 2 min at 70°C, then dehydrated at 4°C. The slides were treated with proteinase K (10 µg/ml in 20 mM Tris-HCl, 2 mM CaCl2) at 37°C for 8-10 min and dehydrated. After 20 preannealing, the hybridization mixture containing the probe was placed on the slide, covered with a coverslip, sealed with rubber cement and incubated overnight in a humid chamber at 37°C. After hybridization and post hybridization washes, the biotinylated probe was detected by avidin-FITC (5 µg/ml, Vector Laboratories) and amplified once with additional layers of biotinylated goat anti-avidin (5 µg/ml, Vector Laboratories) and avidin-FITC. For chromosomal localization, fluorescent R-Bands were obtained as described by Cherif et al. (1990). The slides were observed under a 25 LEICA fluorescent microscope (DMRXA). Chromosomes were counterstained with propidium iodide and the fluorescent signal of the probe appeared as two symmetrical yellow-green spots on both chromatids of the fluorescent R-band chromosome.

Localization

A specific signal (a double yellow-green spot) was observed on band 11q12-q13 on at least 30 on chromosome 11 in >80% of the metaphases with all the probes.

EXAMPLE 2 : IDENTIFICATION OF BIALLELIC MARKERS: DNA EXTRACTION

Donors were unrelated and healthy. They presented a sufficient diversity for being representative of a French heterogeneous population. The DNA from 100 individuals was extracted and tested for the 35 detection of the biallelic markers.

30 ml of peripheral venous blood were taken from each donor in the presence of EDTA. Cells (pellet) were collected after centrifugation for 10 minutes at 2000 rpm. Red cells were lysed by a lysis solution (50 ml final volume : 10 mM Tris pH7.6; 5 mM MgCl₂; 10 mM NaCl). The solution was centrifuged (10 minutes, 2000 rpm) as many times as necessary to eliminate the residual red 5 cells present in the supernatant, after resuspension of the pellet in the lysis solution.

The pellet of white cells was lysed overnight at 42°C with 3.7 ml of lysis solution composed of:

- 3 ml TE 10-2 (Tris-HCl 10 mM, EDTA 2 mM) / NaCl 0.4 M
- 200 µl SDS 10%
- 500 µl K-proteinase (2 mg K-proteinase in TE 10-2 / NaCl 0.4 M).

10 For the extraction of proteins, 1 ml saturated NaCl (6M) (1/3.5 v/v) was added. After vigorous agitation, the solution was centrifuged for 20 minutes at 10000 rpm.

For the precipitation of DNA, 2 to 3 volumes of 100% ethanol were added to the previous supernatant, and the solution was centrifuged for 30 minutes at 2000 rpm. The DNA solution was 15 rinsed three times with 70% ethanol to eliminate salts, and centrifuged for 20 minutes at 2000 rpm. The pellet was dried at 37°C, and resuspended in 1 ml TE 10-1 or 1 ml water. The DNA concentration was evaluated by measuring the OD at 260 nm (1 unit OD = 50 µg/ml DNA).

20 To determine the presence of proteins in the DNA solution, the OD 260 / OD 280 ratio was determined. Only DNA preparations having a OD 260 / OD 280 ratio between 1.8 and 2 were used in the subsequent examples described below.

The pool was constituted by mixing equivalent quantities of DNA from each individual.

EXAMPLE 3 : IDENTIFICATION OF BIALLELIC MARKERS: AMPLIFICATION OF GENOMIC DNA BY PCR

25 The amplification of specific genomic sequences of the DNA samples of example 2 was carried out on the pool of DNA obtained previously. In addition, 50 individual samples were similarly amplified.

PCR assays were performed using the following protocol:

	Final volume	25 µl
	DNA	2 ng/µl
30	MgCl ₂	2 mM
	dNTP (each)	200 µM
	primer (each)	2.9 ng/µl
	Ampli Taq Gold DNA polymerase	0.05 unit/µl
	PCR buffer (10x = 0.1 M TrisHCl pH8.3 0.5M KCl	1x

35 Each pair of first primers was designed using the sequence information of the olfactory receptor gene cluster disclosed herein and the OSP software (Hillier & Green, 1991). This first pair

of primers was about 20 nucleotides in length and had the sequences disclosed in Table 1 in the columns labeled PU and RP.

Table 1

Amplicon	Position range of the amplicon in SEQ ID 1		Primer name RP	Position range of amplification primer in SEQ ID No 1		Primer name PU	Complementary position range of amplification primer in SEQ ID No 1	
99-13670	7362	7824	B1	7362	7380	C1	7805	7824
99-13669	8120	8662	B2	8120	8140	C2	8643	8662
99-13666	14308	14757	B3	14308	14328	C3	14740	14757
99-13664	19346	19845	B4	19346	19366	C4	19826	19845
99-13663	20298	20800	B5	20298	20318	C5	20781	20800
99-13660	76752	77223	B6	76752	76772	C6	77205	77223
99-13652	90967	91494	B7	90967	90987	C7	91474	91494
99-13671	133925	134393	B8	133925	133945	C8	134375	134393
99-13649	139807	140351	B9	139807	139826	C9	140331	140351
99-13648	140912	141434	B10	140912	140932	C10	141416	141434
99-13647	143828	144309	B11	143828	143847	C11	144292	144309

5 Preferably, the primers contained a common oligonucleotide tail upstream of the specific bases targeted for amplification which was useful for sequencing.

Primers PU contain the following additional PU 5' sequence :

TGTAAAACGACGGCCAGT; primers RP contain the following RP 5' sequence :

CAGGAAACAGCTATGACC. The primer containing the additional PU 5' sequence is listed in

10 SEQ ID No 26. The primer containing the additional RP 5' sequence is listed in SEQ ID No 27.

The synthesis of these primers was performed following the phosphoramidite method, on a GENSET UFPS 24.1 synthesizer.

15 DNA amplification was performed on a Genius II thermocycler. After heating at 95°C for 10 min, 40 cycles were performed. Each cycle comprised: 30 sec at 95°C, 54°C for 1 min, and 30 sec at 72°C. For final elongation, 10 min at 72°C ended the amplification. The quantities of the amplification products obtained were determined on 96-well microtiter plates, using a fluorometer and Picogreen as intercalant agent (Molecular Probes).

EXAMPLE 4 : IDENTIFICATION OF BIALLELIC MARKERS: SEQUENCING OF AMPLIFIED GENOMIC DNA AND IDENTIFICATION OF POLYMORPHISMS.

20 The sequencing of the amplified DNA obtained in example 3 was carried out on ABI 377 sequencers. The sequences of the amplification products were determined using automated dideoxy terminator sequencing reactions with a dye terminator cycle sequencing protocol. The products of the sequencing reactions were run on sequencing gels and the sequences were determined using gel image analysis (ABI Prism DNA Sequencing Analysis software (2.1.2 version)).

The sequence data were further evaluated using the above mentioned polymorphism analysis software designed to detect the presence of biallelic markers among the pooled amplified fragments. The polymorphism search was based on the presence of superimposed peaks in the electrophoresis pattern resulting from different bases occurring at the same position as described previously.

5 11 fragments of amplification were analyzed. In these segments, 13 biallelic markers referred to as A1 to A13 in the BM column were detected. The localization of these biallelic markers is as shown in Table 2.

Table 2

Amplicon	BM	Marker Name	Localization in <i>OLF</i> gene cluster	Polymorphism	BM position in SEQ ID No 1
99-13670	A1	99-13670-305	Between Orf1 and Orf2	A/C	7521
99-13669	A2	99-13669-471	Between Orf1 and Orf2	A/C	8192
99-13666	A3	99-13666-275	Between Orf2 and Orf3	A/T	14483
99-13664	A4	99-13664-221	Between Orf2 and Orf3	A/G	19625
99-13663	A5	99-13663-218	Between Orf2 and Orf3	C/T	20583
99-13660	A6	99-13660-277	Between Orf4 and Orf5	G/T	76947
99-13652	A7	99-13652-407	Between Orf5 and Orf6	G/C	91088
99-13652	A8	99-13652-357	Between Orf5 and Orf6	C/T	91138
99-13652	A9	99-13652-308	Between Orf5 and Orf6	C/T	91187
99-13671	A10	99-13671-396	Between Orf9 and Orf10	C/T	133998
99-13649	A11	99-13649-286	Between Orf9 and Orf10	A/G	140066
99-13648	A12	99-13648-259	Between Orf9 and Orf10	C/T	141176
99-13647	A13	99-13647-278	After Orf10	C/T	144033

10

Table 3

BM	Marker Name	Position range of probes in SEQ ID No 1		Probes
A1	99-13670-305	7498	7544	P1
A2	99-13669-471	8169	8215	P2
A3	99-13666-275	14460	14506	P3
A4	99-13664-221	19602	19648	P4
A5	99-13663-218	20560	20606	P5
A6	99-13660-277	76924	76970	P6
A7	99-13652-407	91065	91111	P7
A8	99-13652-357	91115	91161	P8
A9	99-13652-308	91164	91210	P9
A10	99-13671-396	133975	134021	P10
A11	99-13649-286	140043	140089	P11
A12	99-13648-259	141153	141199	P12
A13	99-13647-278	144010	144056	P13

EXAMPLE 5 : VALIDATION OF THE POLYMORPHISMS THROUGH MICROSEQUENCING

The biallelic markers identified in example 4 were further confirmed and their respective frequencies were determined through microsequencing. Microsequencing was carried out for each 5 individual DNA sample described in Example 2.

Amplification from genomic DNA of individuals was performed by PCR as described above for the detection of the biallelic markers with the same set of PCR primers (Table 1).

The preferred primers used in microsequencing were about 19 nucleotides in length and hybridized just upstream of the considered polymorphic base. According to the invention, the 10 primers used in microsequencing are detailed in Table 4.

Table 4

Marker Name	BM	Mis. 1	Position range of microsequencing primer mis 1 in SEQ ID No 1	Mis. 2	Complementary position range of microsequencing primer mis. 2 in SEQ ID No 1
99-13670-305	A1	D1	7502	7520	E1
99-13669-471	A2	D2	8173	8191	E2
99-13666-275	A3	D3	14464	14482	E3
99-13664-221	A4	D4	19606	19624	E4
99-13663-218	A5	D5	20564	20582	E5
99-13660-277	A6	D6	76928	76946	E6
99-13652-407	A7	D7	91069	91087	E7
99-13652-357	A8	D8	91119	91137	E8
99-13652-308	A9	D9	91168	91186	E9
99-13671-396	A10	D10	133979	133997	E10
99-13649-286	A11	D11	140047	140065	E11
99-13648-259	A12	D12	141157	141175	E12
99-13647-278	A13	D13	144014	144032	E13

Mis 1 and Mis 2 respectively refer to microsequencing primers which hybridized with the non-coding strand of the olfactory receptor gene or with the coding strand of the olfactory receptor 15 gene.

The microsequencing reaction was performed as follows :

After purification of the amplification products, the microsequencing reaction mixture was prepared by adding, in a 20 μ l final volume: 10 pmol microsequencing oligonucleotide, 1 U Thermosequenase (Amersham E79000G), 1.25 μ l Thermosequenase buffer (260 mM Tris HCl pH 20 9.5, 65 mM MgCl₂), and the two appropriate fluorescent ddNTPs (Perkin Elmer, Dye Terminator Set 401095) complementary to the nucleotides at the polymorphic site of each biallelic marker tested, following the manufacturer's recommendations. After 4 minutes at 94°C, 20 PCR cycles of 15 sec at 55°C, 5 sec at 72°C, and 10 sec at 94°C were carried out in a Tetrad PTC-225 thermocycler (MJ Research). The unincorporated dye terminators were then removed by ethanol precipitation. Samples 25 were finally resuspended in formamide-EDTA loading buffer and heated for 2 min at 95°C before

being loaded on a polyacrylamide sequencing gel. The data were collected by an ABI PRISM 377 DNA sequencer and processed using the GENESCAN software (Perkin Elmer).

Following gel analysis, data were automatically processed with software that allows the determination of the alleles of biallelic markers present in each amplified fragment.

5 The software evaluates such factors as whether the intensities of the signals resulting from the above microsequencing procedures are weak, normal, or saturated, or whether the signals are ambiguous. In addition, the software identifies significant peaks (according to shape and height criteria). Among the significant peaks, peaks corresponding to the targeted site are identified based on their position. When two significant peaks are detected for the same position, each sample is
10 categorized classification as homozygous or heterozygous type based on the height ratio.

While the preferred embodiment of the invention has been illustrated and described, it will be appreciated that various changes can be made therein by the one skilled in the art without departing from the spirit and scope of the invention.

15

REFERENCES :

Adra et al., 1987, *Gene*, **60**:65-74

Beard et al., 1980, *Virology*, Vol. 75:81

Beaucage et al., *Tetrahedron Lett* 1981, **22**: 1859-1862

Bolmont et al., *J. of Submicroscopic cytology and pathology*, 1990, **22**:117-122

20 Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979; **68**:109-151

Chai H. et al., 1993, *Biotechnol. Appl. Biochem.*, **18**:259-273

Cherif D, Julier C, Delattre O, Derré J, Lathrop GM and Berger R, *P.N.A.S. USA* (1990) **87**: 6639-6643

Compton J. *Nature*. 1991 Mar 7; **350**(6313): 91-92.

25 Current Protocols in Molecular Biology, 1989, Ausubel FM et al. (eds), Greene Publishing Associates, Sections 9.10-9.14.

Davis, L. et al. Basic Methods in Molecular Biology Elsevier, New York. Section 21-2

Engvall E, 1980, *Methods Enzymol*, **70**(A):419-439

Feldman and Steg, 1996, *Medecine/Sciences, synthese*, **12**:47-55

30 Felgner et al., 1987, *Proc. Natl. Acad. Sci.*, **84**:7413

Fisher, D., Chap. 42 in: *Manual of Clinical Immunology*, 2d Ed. (Rose and Friedman, Eds.) Amer. Soc. For Microbiol., Washington, D.C. (1980).

Flotte et al., 1992, *Am. J. Respir. Cell Mol. Biol.*, **7** : 349-356.

Fraley et al., 1980, *J. Biol. Chem.*, **255**:10431).

35 Fuller S.A. et al., 1996, *Immunology in Current Protocols in Molecular Biology*, Ausubel et al. Eds, John Wiley & Sons, Inc., USA

Graham, 1984, EMBO J., **3**:2917

Guatelli JC et al., 1990, Proc. Natl. Acad. Sci. USA, **87** : 1874-1878.

Guzman, 1981, Cell, **23** : 175

Hames BD and Higgins SJ, 1985, "Nucleic acid hybridization : a practical approach", Hames and Higgins Ed., IRL Press, Oxford.

5 Harju L, et al., *Clin Chem* 1993;39(11Pt 1):2282-2287

Hartley JW, Rowe WP, *J Virol* 1976;19(1):19-25

Hillier L. and Green P. *Methods Appl.*, 1991, 1: 124-8.

Houbenweyl, 1974, in Meuthode der Organischen Chemie, E. Wunsch Ed., Volume 15-I et 15-II,

10 Thieme, Stuttgart

Huygen et al., 1996, *Nature Medicine*, **2**(8):893-898

Julan et al., 1992, *J. Gen. Virol.*, **73** : 3251 – 3255.

Kaneda et al., 1989, *Science*, **243**:375

Kiefer H et al., 1996, *Biochemistry*, **35**(50):16077-16084.

15 Koch Y., 1977, *Biochem. Biophys. Res. Commun.*, **74**:488-491

Kohler G. and Milstein C., 1975, *Nature*, **256** : 495.

Kort et al., 1983, *Nucleic Acids Research*, **11**:8287-8301

Lenhard T. et al., 1996, *Gene*, **169**:187-190

Levrero et al., 1991, *Gene*, **101**:195

20 Lin Z, Floros J, 1998, *Biotechniques*, **24**(6):937-940

Livak KJ, and Hainer JW., 1994, *Hum Mutat.*, **3**(4): 379-385.

Mackey K, Steinkamp A, Chomczynski P, 1998, *Mol Biotechnol*, **9**(1):1-5

Mansour SL, Thomas KR, Capecchi MR, *Nature* 1988;336(6197):348-52

McLaughlin et al., 1989, *J. Virol.*, **62** : 1963 – 1973.

25 Merrifield RB 1965 *Nature*. **207**: 522-523.

Merrifield RB 1965 *Science*. **150**: 178-185.

Midoux, 1993, *Nucleic Acids Research*, **21**:871-878

Muzyczka et al., 1992, *Curr. Topics in Micro. and Immunol.*, **158** : 97-129.

Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;68:90-98

30 Neda et al., 1991, *J. Biol. Chem.*, **266** : 14143 – 14146.

Nekrasova E, Sosinskaya A, Natochin M, Lancet D, Gat U, *Eur J Biochem* 1996;238(1):28-37

O'Reilly et al., 1992, *Baculovirus expression vectors : a Laboratory Manual*. W.H. Freeman and Co., New York

Adra et al., 1987, *Gene*, **60**:65-74

Ohno et al., 1994, *Sciences*, **265**:781-784

35 Ohwada A, Hirschowitz EA, Crystal RG, *Hum Gene Ther* 1996;7:1567-76

Ouchterlony, O. et al., Chap. 19 in: *Handbook of Experimental Immunology* D. Wier (ed) Blackwell (1973)

Pagano et al., 1967, *J. Virol.*, **1**:891

Palmer S et al., 1989, *Biochim Biophys Acta*, **1014**(3):239-246.

Pastore, 1994, *Circulation*, **90**:I-517

PCR Methods and Applications" (1991, Cold Spring Harbor Laboratory Press

5 Pearson WR, Lipman DJ, *Proc Natl Acad Sci U S A* 1988;85(8):2444-8

Porath J et al., 1975, *Nature*, **258**(5536) : 598-599.

Raming K et al., 1993, *Nature*, **361**(6410):353-356

Rettlez and Basenga, 1987, *Mol. Cell. Biol.*, **7**:1676-1685

Romana SP, Tachdjian G, Druart L, Cohen D, Berger R, Cherif D *Eur J Hum Genet* (1993) **1**: 245-10 251.

Roth J.A. et al., 1996, *Nature Medicine*, **2**(9):985-991

Rougeot, C. et al., *Eur. J. Biochem.* **219** (3): 765-773, 1994

Sambrook, J. Fritsch, E. F., and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York

15 Samulski et al., 1989, *J. Virol.*, **63** : 3822-3828.

Sanchez-Pescador R., 1988, *J. Clin. Microbiol.*, **26**(10):1934-1938

Sczakiel G. et al., 1995, *Trends Microbiol.*, 1995, **3**(6):213-217

Smith et al., 1983, *Mol. Cell. Biol.*, **3**:2156-2165.

Steiner AL et al., 1972, *J Biol Chem*, **247**(4):1106-1113

20 Syvanen AC, et al., 1994, *Hum Mutat.*, **3**(3): 172-179.

Tacson et al., 1996, *Nature Medicine*, **2**(8):888-892

Tucker J, Grisshammer R, *Biochem J* 1996;317(Pt 3):891-899

Urdea M.S., 1988, *Nucleic Acids Research*, **11**: 4937-4957

Urdea MS et al., 1991, *Nucleic Acids Symp Ser.*, **24**: 197-200.

25 Vaitukaitis J. et al., 1971, *J. Clin. Endocrinol. Metab.*, **33** : 988-991

Vlasak R. et al., 1983, *Eur. J. Biochem.*, **135**:123-126

White, B.A. Molecular Cloning to Genetic Engineering Ed. in Methods in Molecular Biology **67**: Humana Press, Totowa 1997

Zhao H et al., 1998, *Science*, **279**(5348):237-242

30 Zhenlin et al., *Gene*, 1989, **78**:243-254.

SEQUENCE LISTING FREE TEXT

The following free text appears in the accompanying Sequence Listing :

open reading frame

ubiquitin 1 pseudogene complement

5 ubiquitin 2 pseudogene complement

polymorphic base

or

complement

probe

10 sequencing oligonucleotide PrimerPU

 sequencing oligonucleotide PrimerRP

What is claimed:

1. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of SEQ ID No 1 or the complements thereof, wherein said contiguous span comprises at least 1 of the following nucleotide positions of SEQ ID No 1: 1-113643, 114064-5 127488, 127855-144460.
2. An isolated, purified, or recombinant polynucleotide comprising a contiguous span of at least 12 nucleotides of a sequence selected from the group consisting of SEQ ID Nos 2-11 or the complements thereof.
- 10 3. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 or the complement thereof, wherein said span includes an olfactory receptor-related biallelic marker in said sequence.
- 15 4. A polynucleotide according to claim 3, wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13, and the complements thereof.
- 20 5. A polynucleotide according to claims 3 or 4, wherein said contiguous span is 18 to 47 nucleotides in length and said biallelic marker is within 4 nucleotides of the center of said polynucleotide.
- 25 6. A polynucleotide according to claim 5, wherein said polynucleotide consists essentially of a sequence selected from the following sequences: P1 to P13, and the complementary sequences thereto.
7. A polynucleotide according to any one of claims 1, 2 or 3, wherein the 3' end of said contiguous span is present at the 3' end of said polynucleotide.
- 30 8. A polynucleotide according to claims 3 or 4, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide and said biallelic marker is present at the 3' end of said polynucleotide.
- 35 9. An isolated, purified, or recombinant polynucleotide consisting essentially of a contiguous span of 8 to 50 nucleotides of SEQ ID No 1 or the complement thereof, wherein the 3' end of said contiguous span is located at the 3' end of said polynucleotide, and wherein the 3' end of said

polynucleotide is located within 20 nucleotides upstream of an olfactory receptor-related biallelic marker in said sequence.

10. A polynucleotide according to claim 9, wherein the 3' end of said polynucleotide is
5 located 1 nucleotide upstream of said olfactory receptor-related biallelic marker in said sequence.

11. A polynucleotide according to claim 10, wherein said polynucleotide consists
essentially of a sequence selected from the following sequences: D1 to D13, and E1 to E13.

10 12. A polynucleotide according to claim 7 consisting essentially of a sequence selected from
the following sequences: B1 to B11 and C1 to C11.

13. An isolated, purified, or recombinant polynucleotide which encodes a polypeptide
comprising a contiguous span of at least 6 amino acids of a sequence selected from the group
15 consisting of SEQ ID Nos 12-21.

14. A polynucleotide for use in a genotyping assay for determining the identity of the
nucleotide at an olfactory receptor-related biallelic marker or the complement thereof.

20 15. A polynucleotide according to claim 14, wherein the polynucleotide is used in an assay
selected from the group consisting of: a hybridization assay, a sequencing assay, an enzyme-based
mismatch detection assay, and an amplification of a segment of nucleotides comprising said biallelic
marker.

25 16. A polynucleotide according to any one of claims 1-15 attached to a solid support.

17. An array of polynucleotides comprising at least one polynucleotide according to claim
16.

30 18. An array according to claim 17, wherein said array is addressable.

19. A polynucleotide according to any one of claims 1-15, further comprising a label.

20. A recombinant vector comprising a polynucleotide according to any one of claims 1-15.

35

21. A host cell comprising a recombinant vector according to claim 20.

22. A non-human host animal or mammal comprising a recombinant vector according to claim 20.

23. A mammalian host cell comprising an olfactory receptor gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1-15.

24. A non-human host mammal comprising an olfactory receptor gene disrupted by homologous recombination with a knock out vector, comprising a polynucleotide according to any one of claims 1-15.

25. An isolated, purified, or recombinant polypeptide comprising a contiguous span of at least 6 amino acids of a sequence selected from the group consisting of SEQ ID Nos 12-21.

15 26. An isolated or purified antibody composition are capable of selectively binding to an epitope-containing fragment of a polypeptide according to claim 25.

27. A method of genotyping comprising determining the identity of a nucleotide at an olfactory receptor-related biallelic marker or the complement thereof in a biological sample.

20 28. A method according to claim 27, wherein said biological sample is derived from a single subject.

25 29. A method according to claim 28, wherein the identity of the nucleotides at said biallelic marker is determined for both copies of said biallelic marker present in said individual's genome.

30 30. A method according to claim 27, wherein said biological sample is derived from multiple subjects.

30 31. A method according to claim 27, further comprising amplifying a portion of said sequence comprising the biallelic marker prior to said determining step.

32. A method according to claim 31, wherein said amplifying step is performed by PCR.

35 33. A method according to claim 27, wherein said determining is performed by an assay selected from the group consisting of: a hybridization assay, a sequencing assay, a microsequencing assay, and an enzyme-based mismatch detection assay.

34. A method according to claim 27 wherein said olfactory receptor-related biallelic marker is selected from the group consisting of A1 to A13 and the complements thereof.

5 35. A method for the screening of a candidate substance interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, or fragments or variants thereof, comprises the following steps :

- a) providing a polypeptide selected from the group consisting of the sequences of SEQ ID Nos 12-21 , or a peptide fragment or a variant thereof;
- 10 b) obtaining a candidate substance;
- c) bringing into contact said polypeptide with said candidate substance; and
- d) detecting the complexes formed between said polypeptide and said candidate substance.

15 36. A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, wherein said method comprises :

- a) providing a recombinant eukaryotic host cell containing a nucleic acid encoding a polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21;
- 20 b) preparing membrane extracts of said recombinant eukaryotic host cell;
- c) bringing into contact the membrane extracts prepared at step b) with a selected ligand molecule; and
- d) detecting the production level of second messengers metabolites.

25 37. A method for the screening of ligand molecules interacting with an olfactory receptor polypeptide selected from the group consisting of SEQ ID Nos 12-21, wherein said method comprises :

- a) providing an adenovirus containing a nucleic acid encoding a polypeptide selected from the group consisting of the polypeptides comprising the amino acid sequences SEQ ID Nos 12-21;
- 30 b) infecting an olfactory epithelium with said adenovirus;
- c) bringing into contact the olfactory epithelium b) with a selected ligand molecule; and
- d) detecting the increase of the response to said ligand molecule.

FIGURE 1

TM1

	1	50
list1.msf{orf-8}	~MRRNFTLVT EFILLGLTNH	QELQILLFML FLAIYMVTVA GNLSMIALIQ
list1.msf{orf-9}	~MRRNCTLVT EFILLGLTSR	RELQILLFTL FLAIYMVTVA GNLMIVLHQ
list1.msf{orf-7}	-----	-----
list1.msf{orf-2}	MFSPNHTIVT EFILLGLTDD	PVLEKILFGV FLAIYLITLA GNLCMILLIR
list1.msf{orf-4}	MLSPNHTIVT EFILLGLTDD	PVLEKILFGV FLAIYLITLA GNLCMILLIR
list1.msf{orf-5}	MSNTNGSAIT EFILLGLTDC	PELQSLLFVL FLVYVLVTLL GNLMIMLMR
list1.msf{orf-6}	MVRGNSTLVT EFILLGLKDL	PELQPILFVL FLLIYLITVG GNLMGLVLIR
list1.msf{orf-10}	MSRRNYTELT EFVLLGLTSR	PELRAVAFAL FLFVYIATVV GNLMIIILIK
list1.msf{orf-3}	MLKKNHTAVT EFVLLGLTDR	AELQSLLFVV FLVIYLITVI GNVSMILLIR
list1.msf{orf-1}	-----	-----
Consensus	M-R-N-T-VT EFILLGLTD-	PELQ-LLF-L FLAIYLITVA GNLMGI-LIR

TM2

	51	100
list1.msf{orf-8}	ANARLHTPMY FFLSHLSFLD	LCFSSNVTPK MLEIFLSEKK SISYPACLVQ
list1.msf{orf-9}	ANAWLHMPMY FFLSHLSFVD	LCFSSNVTPK MLEIFLSEKK SISYPACLVQ
list1.msf{orf-7}	PTPRLHTPMY FFLSNLSFVD	LCFSSNVTPR MLEIFLSEKK SISYPARLVQ
list1.msf{orf-2}	TNSHLQTPMY FFLGHLSFVD	ICYSSNVTPN MLHNFLSEQK TISYAGCFTQ
list1.msf{orf-4}	TNSQLQTPMY FFLGHLSFLD	ICYSSNVTXN MLHNFLSEQK TISYAGCFTQ
list1.msf{orf-5}	LDSRLHTPMY FFLTNLAFVD	LCYTSNATPQ MSTNIVSE.K TISFAGCFTQ
list1.msf{orf-6}	IDSRRLHTPMY FFLASLSCLD	LYYSTNVTPK MLVNFFSDKK AISYAACLVQ
list1.msf{orf-10}	VDSRLHTPMY-----	-----
list1.msf{orf-3}	SDSTLHTPMY FFLSHLSFVD	LCYTTNVTPQ MLVNFLSKRK TISFIGCFIQ
list1.msf{orf-1}	SDSTLHTPMC LFLSHLSFVD	LYYATNATPP MLVNFFFPRE KPFPLLVALS
Consensus	-DSRLHTPMY FFLSHLSFVD	LCYSSNVTP- ML-NFLSEKK TISYA-C-VQ

TM3

	101	150
list1.msf{orf-8}	CYLYIILVHV EIYILAVMAF	D-----
list1.msf{orf-9}	CYLFIALVHV EIYILAVMAF	DRYMAICNPL LYGSRMSKSV CSFLITVPYV
list1.msf{orf-7}	CYLFITLVHV ELYILAVMAF	DRYMAICNPL LYGSRMSKSV CSFLITVLYV
list1.msf{orf-2}	CLLFIALVIT EFYFLASMAL	DRYVAICSPY HYSSRMSKNI CISLVTVPYM
list1.msf{orf-4}	CLLFIALVIT EFYFLASMAL	DRYVAICSPY HYSSRMSKNI CISLVTVPYM
list1.msf{orf-5}	CYIFIALLT EFYMLAAMAY	DRYVAIYDPL RYSVKTTSRRV CICLATFPYV
list1.msf{orf-6}	CYFFIAVVT EYYMLAVMAY	DRYVAICNPL LYSSKMSKGL CIRLIAGPYV
list1.msf{orf-10}	-----	-----
list1.msf{orf-3}	FHFFIALVIT DYYMLTVMAY	DRYMAICKPL LYGSKMTROV CLCLAAAPYI
list1.msf{orf-1}	NFTFSLHW~	-----
Consensus	CYLFIALVIT E-Y-LAVMA-	DRYVAIC-PL LYSSRMSK-V CI-L-TV PYV

TM4

	151	200
list1.msf{orf-8}	-----	-----
list1.msf{orf-9}	YGALTGLMET MWTYNLAFCG	PNEINHFYCA DPPLIKLACS DTYNKELSMF
list1.msf{orf-7}	YGALTGLMET MWTYNLAFCG	PSEINHFYCV DPPLIKLACS DTYNKEVSMF
list1.msf{orf-2}	YGFLNGLSQT LLTFHLSFCG	SLEINHFYCA DPPLIMLACS DTRVKKMAMF
list1.msf{orf-4}	YGXLNGLSQT LLTFHLSFCG	SLEINHFYCA DPPLIMLACS DTRVKKMAMF
list1.msf{orf-5}	YGFSDGLFQA ILTFRLTFCR	SNVINHFYCA DPPLIKLSCS DTYVKEHAMF
list1.msf{orf-6}	YGFLSGLMET MWTYHLTFCG	SNIINHFYCA DPPLIRLSCS DTFIKEHSMF
list1.msf{orf-10}	-----	-----
list1.msf{orf-3}	YGFANGLSTD HPDASSVLLW	TQ-----
list1.msf{orf-1}	-----	-----
Consensus	YGFL-GL--T --T--L-FCG	S-EINHFYCA DPPLI-LACS DT--KE--MF

FIGURE 1 (continued)

	TM5	TM6
	201	250
list1.msf{orf-8}	-----	-----
list1.msf{orf-9}	IVAGWNLSFS LFIICISYLY	IFPAILKIRS TEGRQKAFST CGSHLTAVTI
list1.msf{orf-7}	VVAGFNFTYP LLIILISYLY	IFPATLRICS TEGRHKAFT CGSHLTAVTI
list1.msf{orf-2}	VVAGFTLSSS LFIILLSYLF	IFAAIFRIRS AEGRHKAFT CASHLTIVTL
list1.msf{orf-4}	VVAGFTLSSS LFIILLSYLF	IFAAIFRIRS AEGRHKAFT CASHLTIVTL
list1.msf{orf-5}	ISAGFNLSSS LTIVLVSYAF	ILAAILRIKS AEGRHKAFT CGSHMMAVTL
list1.msf{orf-6}	VVA-----	-----
list1.msf{orf-10}	-----	-----
list1.msf{orf-3}	-----	-----
list1.msf{orf-1}	-----	-----
Consensus	VVAGF-LS-S L-IIL-SYL	IF-AI-RI-S -EGRHKAFST C-SHLT-VT-

	TM6	TM7	
	251	300	
list1.msf{orf-8}	-----	-----	-----
list1.msf{orf-9}	FYATLFFMYL	RPPSKESVEQ GKMVAVFYTT VIPMLNLIIY SLRNKNVKEA	
list1.msf{orf-7}	FYSALFFMYL	RRPSEESMEQ GKMVAVFYTT VIPMLNPMIY SLRNKDVKEA	
list1.msf{orf-2}	FYGTLCFCMYV	RPPSEKSVEE SKITAVFYTF LTPMLNPLIY SLRNTDVILA	
list1.msf{orf-4}	FYGTLCFCMYV	RPPSEKSVEE SKIIIAVFYTF LSPMLNPLIY SLRNRDVILA	
list1.msf{orf-5}	FYGTLCFCMYI	RPPTDKTVEE SKIIIAVFYTF VSPVLNPLIY SLRNKDVQQA	
list1.msf{orf-6}	-----	-----	-----
list1.msf{orf-10}	-----	-----	-----
list1.msf{orf-3}	-----	-----	-----
list1.msf{orf-1}	-----	-----	-----
Consensus	FY-TLF-MY-	RPPS--SVE- -K--AVFYT- --PMLNP-IY	SLRN-DV--A

	301	315
list1.msf{orf-8}	-----	-----
list1.msf{orf-9}	LIKELSMKIY FS---	
list1.msf{orf-7}	LCKELFKRKL FSK--	
list1.msf{orf-2}	MQQMIRGKSF HKIAV	
list1.msf{orf-4}	IQQMIRGKSF CKIAV	
list1.msf{orf-5}	LKNVLR-----	
list1.msf{orf-6}	-----	
list1.msf{orf-10}	-----	
list1.msf{orf-3}	-----	
list1.msf{orf-1}	-----	
Consensus	-----	

<110> Genset SA

<120> Genes encoding olfactory receptors and biallelic markers thereof.

<150> US 60/104,299

<151> 1999-10-13

<160> 27

<170> Patent.pm

<210> 1

<211> 144460

<212> DNA

<213> Homo sapiens

<220>

<221> CDS

<222> 2406..2600

<223> open reading frame 1

<220>

<221> CDS

<222> 9711..10658

<223> open reading frame 2

<220>

<221> CDS

<222> 24851..25369

<223> open reading frame 3

<220>

<221> CDS

<222> 45714..46661

<223> open reading frame 4

<220>

<221> CDS

<222> 80198..81115

<223> open reading frame 5

<220>

<221> CDS

<222> 96291..96902

<223> open reading frame 6

<220>

<221> CDS

<222> 110758..111564

<223> open reading frame 7

<220>

<221> CDS

<222> 122525..122887

<223> open reading frame 8

<220>

<221> CDS

<222> 132454..133389

<223> open reading frame 9

<220>

```
<221> CDS
<222> 143398..143577
<223> open reading frame 10

<220>
<221> misc_feature
<222> 113644..114063
<223> ubiquitin 1 pseudogene complement

<220>
<221> misc_feature
<222> 127489..127854
<223> ubiquitin 2 pseudogene complement

<220>
<221> allele
<222> 7521
<223> 99-13670-305 : polymorphic base G or T

<220>
<221> allele
<222> 8192
<223> 99-13669-471 : polymorphic base G or T

<220>
<221> allele
<222> 14483
<223> 99-13666-275 : polymorphic base A or T

<220>
<221> allele
<222> 19625
<223> 99-13664-221 : polymorphic base C or T

<220>
<221> allele
<222> 20583
<223> 99-13663-218 : polymorphic base A or G

<220>
<221> allele
<222> 76947
<223> 99-13660-277 : polymorphic base A or C

<220>
<221> allele
<222> 91088
<223> 99-13652-407 : polymorphic base G or C

<220>
<221> allele
<222> 91138
<223> 99-13652-357 : polymorphic base A or G

<220>
<221> allele
<222> 91187
<223> 99-13652-308 : polymorphic base A or G

<220>
<221> allele
<222> 133998
<223> 99-13671-396 : polymorphic base A or G
```

```
<220>
<221> allele
<222> 140066
<223> 99-13649-286 : polymorphic base C or T

<220>
<221> allele
<222> 141176
<223> 99-13648-259 : polymorphic base A or G

<220>
<221> allele
<222> 144033
<223> 99-13647-278 : polymorphic base A or G

<220>
<221> primer_bind
<222> 7362..7380
<223> 99-13670.rp

<220>
<221> primer_bind
<222> 7805..7824
<223> 99-13670.pu complement

<220>
<221> primer_bind
<222> 8120..8140
<223> 99-13669.rp

<220>
<221> primer_bind
<222> 8643..8662
<223> 99-13669.pu complement

<220>
<221> primer_bind
<222> 14308..14328
<223> 99-13666.rp

<220>
<221> primer_bind
<222> 14740..14757
<223> 99-13666.pu complement

<220>
<221> primer_bind
<222> 19346..19366
<223> 99-13664.rp

<220>
<221> primer_bind
<222> 19826..19845
<223> 99-13664.pu complement

<220>
<221> primer_bind
<222> 20298..20318
<223> 99-13663.rp

<220>
<221> primer_bind
```

```
<222> 20781..20800
<223> 99-13663.pu complement

<220>
<221> primer_bind
<222> 76752..76772
<223> 99-13660.rp

<220>
<221> primer_bind
<222> 77205..77223
<223> 99-13660.pu complement

<220>
<221> primer_bind
<222> 90967..90987
<223> 99-13652.rp

<220>
<221> primer_bind
<222> 91474..91494
<223> 99-13652.pu complement

<220>
<221> primer_bind
<222> 133925..133945
<223> 99-13671.rp

<220>
<221> primer_bind
<222> 134375..134393
<223> 99-13671.pu complement

<220>
<221> primer_bind
<222> 139807..139826
<223> 99-13649.rp

<220>
<221> primer_bind
<222> 140331..140351
<223> 99-13649.pu complement

<220>
<221> primer_bind
<222> 140912..140932
<223> 99-13648.rp

<220>
<221> primer_bind
<222> 141416..141434
<223> 99-13648.pu complement

<220>
<221> primer_bind
<222> 143828..143847
<223> 99-13647.rp

<220>
<221> primer_bind
<222> 144292..144309
<223> 99-13647.pu complement
```

```
<220>
<221> misc_binding
<222> 7498..7544
<223> 99-13670-305.probe

<220>
<221> misc_binding
<222> 8169..8215
<223> 99-13669-471.probe

<220>
<221> misc_binding
<222> 14460..14506
<223> 99-13666-275.probe

<220>
<221> misc_binding
<222> 19602..19648
<223> 99-13664-221.probe

<220>
<221> misc_binding
<222> 20560..20606
<223> 99-13663-218.probe

<220>
<221> misc_binding
<222> 76924..76970
<223> 99-13660-277.probe

<220>
<221> misc_binding
<222> 91065..91111
<223> 99-13652-407.probe

<220>
<221> misc_binding
<222> 91115..91161
<223> 99-13652-357.probe

<220>
<221> misc_binding
<222> 91164..91210
<223> 99-13652-308.probe

<220>
<221> misc_binding
<222> 133975..134021
<223> 99-13671-396.probe

<220>
<221> misc_binding
<222> 140043..140089
<223> 99-13649-286.probe

<220>
<221> misc_binding
<222> 141153..141199
<223> 99-13648-259.probe

<220>
<221> misc_binding
<222> 144010..144056
```

```
<223> 99-13647-278.probe

<220>
<221> primer_bind
<222> 7502..7520
<223> 99-13670-305.mis

<220>
<221> primer_bind
<222> 7522..7540
<223> 99-13670-305.mis complement

<220>
<221> primer_bind
<222> 8173..8191
<223> 99-13669-471.mis

<220>
<221> primer_bind
<222> 8193..8211
<223> 99-13669-471.mis complement

<220>
<221> primer_bind
<222> 14464..14482
<223> 99-13666-275.mis

<220>
<221> primer_bind
<222> 14484..14502
<223> 99-13666-275.mis complement

<220>
<221> primer_bind
<222> 19606..19624
<223> 99-13664-221.mis

<220>
<221> primer_bind
<222> 19626..19644
<223> 99-13664-221.mis complement

<220>
<221> primer_bind
<222> 20564..20582
<223> 99-13663-218.mis

<220>
<221> primer_bind
<222> 20584..20602
<223> 99-13663-218.mis complement

<220>
<221> primer_bind
<222> 76928..76946
<223> 99-13660-277.mis

<220>
<221> primer_bind
<222> 76948..76966
<223> 99-13660-277.mis complement

<220>
```

```
<221> primer_bind
<222> 91069..91087
<223> 99-13652-407.mis

<220>
<221> primer_bind
<222> 91089..91107
<223> 99-13652-407.mis complement

<220>
<221> primer_bind
<222> 91119..91137
<223> 99-13652-357.mis

<220>
<221> primer_bind
<222> 91139..91157
<223> 99-13652-357.mis complement

<220>
<221> primer_bind
<222> 91168..91186
<223> 99-13652-308.mis

<220>
<221> primer_bind
<222> 91188..91206
<223> 99-13652-308.mis complement

<220>
<221> primer_bind
<222> 133979..133997
<223> 99-13671-396.mis

<220>
<221> primer_bind
<222> 133999..134017
<223> 99-13671-396.mis complement

<220>
<221> primer_bind
<222> 140047..140065
<223> 99-13649-286.mis

<220>
<221> primer_bind
<222> 140067..140085
<223> 99-13649-286.mis complement

<220>
<221> primer_bind
<222> 141157..141175
<223> 99-13648-259.mis

<220>
<221> primer_bind
<222> 141177..141195
<223> 99-13648-259.mis complement

<220>
<221> primer_bind
<222> 144014..144032
<223> 99-13647-278.mis
```

<220>
 <221> primer_bind
 <222> 144034..144052
 <223> 99-13647-278.mis complement

<400> 1
 caattaaagt tttgttcaact ataagtcttt tttggaaaag agagagaaaac attcaaatta 60
 tttacataacc agtttccatt agcatgtgaa gaacaaacag aaacactttt cagggtgaac 120
 aaaattcctg ctacagttat aaaatcctgc atatacttt tactttgtga ttctgaaaaaa 180
 caccgttcta cctggtttat tgaaatgtgt gaaagctcta atgcaatgtt attttttaca 240
 ttttgttaaca cttaagtcat aaagccaagc tattctaaa ccttgcataa acatgttgg 300
 agaaaattatg ttttagtgtt tggtgaaaac attatgtttc gtcacttaag gtgataaatt 360
 gtactcatta aagaactttg aaagttcaca catagccaat ggtttaaaat gcactaatt 420
 agattcccaa ttctcacaaa ggccagttac tctggaccat tcaatcgcca aagaggaaaa 480
 ctggggccat tcccatcccg ggatatggga agtccccgag cttccagcct ggtccttgc 540
 gcccggaaaat ggcattttt gctttgcct ttggatcctg ttcgtgccgc caaaaatgtt 600
 ctgtgtgggg aaagtgcgag gggagagaaa agacacggac atgatacgtt taagggtaaa 660
 caacgtttat cccatgtaa tggccatgca gatatacgaa gcaaatgata taataataag 720
 caaatgatataat aataagcaga ttgatataat aagtagattt caatggaaacg gggaaaagg 780
 aaaatacatc tacattcacc agactatggaa ggattcaaca acagactggg acgcaacagc 840
 ctgggctcca gagtcagata ggttaggcaaa gagatcctag ttctatacag atacgtacca 900
 tggagcagg tcaacttcct aagcacattt agttgtgata aaaatagatg agtttcaagg 960
 gctgatacat tacatgccac actcaagtt gtgttgcattt acaatttca ttgttgcatt 1020
 aatttcaaat aaaagcaatg ttacaacca tgggttcaag agaagtcataa gtgaacacat 1080
 ataataaaaga cttgcaaaaat aataaaagat aaggtctttt aactatcaaa agacttgcag 1140
 aaaagaacca cagaaaaacca ttttaatattt aactgccttc gtatgtaaaa aattctacat 1200
 tattttgtat gttaaaacat caatctcatg cttactaggc tatttcttaa tgacacatgt 1260
 atttacaaat ttgagagaag aggaagaaat atcaggtgac accactgggt taatgcataa 1320
 atgacaaacc taaatgcatt ttaatttctt tttctttaaa tcgagctgag cttcagcccc 1380
 ttcttttgtt ggtgttctt gtcatctacc ttatcacagt aatcacaatg taagcatgat 1440
 cttctttttt ttttttaagt gcacaatattt ttaactgtt aacaatatac ctattgttac 1500
 ctatgggcac aatgatatac agcatatctc tagaattttt tcttgcaaaa ctataactt 1560
 atacctgctg aacagcaaca ccccatattt ccccttcctc cagccgctgc aaccaccttc 1620
 tattctctgt ttctatgagt ttgactattt tggattcctc atataaattt aatcatgcag 1680
 tatttgcctt tccgtgcctg gcttatttca cttaacataa tgcctccag gttcatcata 1740
 tgacaggatt tcttctttt cttaatgtt aataatattt cattacatgt gtgtactaca 1800
 ttttcttcat ctttcaatgg acatttagg tttttctata tctggactat tgtaaataat 1860
 ggtgcaatga acataagat acctatgtct cttcaagagc ttgattttaa ttcttttgg 1920
 tatatgcccga aagtgcaat tgctggttt tatttataattt cgtttttaa ttatttgaag 1980
 actcatcata ctgtttttt tagtggctgc acaattttt attcccacca atgttgcata 2040
 agggttccaa tttcttcata tgcaccaat attgttgc ttttggattt ttttaaaaata 2100
 aagtaacagc catcataaca aatgtgatattt catgttttgc tttcatatgc attttcctga 2160
 tgatttagt gttgagcacc ttttcatattt tatttattt tttatttata ctctaaatgtt 2220
 tgggatacat ctgcagaacaca cgcagggtt gttacatagg atacatgtgc catgggttt 2280
 tgctgcgccc attaacctgt catctacattt aggtatttct gctaatactt cccctcccc 2340
 agccccccgac cccctgacag gtcctgggt gtgtatgtt cttccctgtg tccatgttg 2400
 tgagcatgag cttcttaata agaagtgtt caacactaca cactccaatg tgcttgcct 2460
 tcagtcatct cttctttgtt gatctctattt atgcccacca tgccactcct ccgatgtgg 2520
 ttaactttt ttttcaaga gaaaaacccgt ttcttttattt ggttgcattt tccaaatttca 2580
 ctttttattt gcaactggta tcacagatta tcatatgtt acagtgtatgg tttatgcata 2640
 ctacatggcc atctgcacgc ctttgcattt tggaaagcaaa atgtccagggt gtgtctgc 2700
 ctgtctact gctgtccctt atattttatgg ctctgcataat ggtctggatc aggtcatcct 2760
 gatgtttgtt ctgttcttctt gtaacccaa tgatgcacac cactttttt ttttttggag 2820
 aaaaatgcattt atatgcacat ttaattccac tataaaatttt tgaatggacg gttggagagg 2880
 aaggggagaaa tacatattaa cggagagaat accacccaga aagtatatac aatggggagaa 2940
 aggaacctgt tgatccaatg ttccatattt tttatatttgc atataaggatc atgattattt 3000
 tctcagtagt aagcatctcc cagggtgcac tctgtatgtt aattggatgatc caaccactt 3060
 tattatgcag aaccacccctt cttagtcctc gcctgcattt atacttgcattt ctttatttcc 3120
 gccatgttca tggggctgg tttcaacccctt atctgccttc tcactatcat ctttatttcc 3180
 tacactttca tcttcacaga cattctgcattt atctgcactt cttggggaaatgttgc 3240
 ttctccacccctt gggggccctt ttttttttttgc gtcactgttgc ttcaagggaaatgttgc 3300
 atgtgcctga gggcccccatttgc tggggatcattt gtagaaacagg gggaaaattgttgc 3360

tatatatttt	tgagtcctac	gttaaacc	ttgatctacc	gtctgaggaa	taaaaatgtt	3420
aaaagaacaa	taagggaa	gttat	aaactgttt	ctaagtaagg	tagatatttt	3480
agttgcagg	tatgt	attat	tcttaccaat	taacgagcat	tataaattaa	3540
caaattca	tctgt	cattt	agtgtttt	gtctttgt	acttgcatat	3600
agtgtatacc	aaattt	ttt	ctagagtt	cagtgtcatc	tcagtgaatt	3660
tcatagaaat	ttaaata	gactt	atgtaaa	caataaagaa	cagtgattcc	3720
ttctttagt	ctcatat	tttgc	actgttac	tgtctatgt	gaaaaggaa	3780
tcattacatt	tttata	tttgc	tttgc	ttcttagacac	cttcaaaca	3840
tagtatatg	agtttctcc	tttctt	tttgc	acagataatg	caacaataaa	3900
gtagggaaaa	gagagatc	actgttac	tgtctatgt	gaaaaggaa	gcataagaa	3960
cttcatttt	acttgta	ttttgc	tttgc	agatgtgtt	aatctgtac	4020
tttgc	ccttgag	ttttgc	tttgc	gtgtgtatg	aatctgtac	4080
tagggctgt	caggatgt	tttgc	tttgc	aatgtatac	aggcagtatg	4140
gtcatcgcc	ttctcc	tttgc	tttgc	tttgc	tttgc	4200
ggacctct	cctggaa	tttgc	tttgc	tttgc	tttgc	4260
tatggctcg	tggtatgg	tttgc	tttgc	tttgc	tttgc	4320
tgtgtcgg	aggattag	tttgc	tttgc	tttgc	tttgc	4380
tctgtctc	ggctgccc	tttgc	tttgc	tttgc	tttgc	4440
cttctattt	gagatagg	tttgc	tttgc	tttgc	tttgc	4500
aatgtgctt	tggtattc	tttgc	tttgc	tttgc	tttgc	4560
ggcttacgt	cacatcc	tttgc	tttgc	tttgc	tttgc	4620
tttgc	caca	tttgc	tttgc	tttgc	tttgc	4680
tcctcttgc	gagatag	tttgc	tttgc	tttgc	tttgc	4740
tgccgatg	ggtcttcc	tttgc	tttgc	tttgc	tttgc	4800
atacttgc	tctgt	tttgc	tttgc	tttgc	tttgc	4860
cccacagg	tttgc	tttgc	tttgc	tttgc	tttgc	4920
tgagatac	tttgc	tttgc	tttgc	tttgc	tttgc	4980
aagcaatgc	aaatttcc	tttgc	tttgc	tttgc	tttgc	5040
gcattcac	caatatgt	tttgc	tttgc	tttgc	tttgc	5100
tttcaagt	gtctt	tttgc	tttgc	tttgc	tttgc	5160
cacatgt	tttgc	tttgc	tttgc	tttgc	tttgc	5220
gtgtggc	ccatttct	tttgc	tttgc	tttgc	tttgc	5280
cacatgtt	ggcagatg	tttgc	tttgc	tttgc	tttgc	5340
agatagct	tcagtag	tttgc	tttgc	tttgc	tttgc	5400
ctacctt	accagca	tttgc	tttgc	tttgc	tttgc	5460
caccataa	acacatgc	tttgc	tttgc	tttgc	tttgc	5520
tggaa	ctaaatg	tttgc	tttgc	tttgc	tttgc	5580
catcatgg	tgctatgc	tttgc	tttgc	tttgc	tttgc	5640
acggatga	ctggagg	tttgc	tttgc	tttgc	tttgc	5700
ctgcaagt	tcacttac	tttgc	tttgc	tttgc	tttgc	5760
gaacaac	caccagg	tttgc	tttgc	tttgc	tttgc	5820
ctctgc	caggact	tttgc	tttgc	tttgc	tttgc	5880
cctgtg	ggaatttac	tttgc	tttgc	tttgc	tttgc	5940
aaagt	aaagaaat	tttgc	tttgc	tttgc	tttgc	6000
cttaat	tgatgtt	tttgc	tttgc	tttgc	tttgc	6060
gacagt	tttac	tttgc	tttgc	tttgc	tttgc	6120
gcttcaaa	aatagtaa	tttgc	tttgc	tttgc	tttgc	6180
aata	catcgt	tttgc	tttgc	tttgc	tttgc	6240
taggctgg	tttgataa	tttgc	tttgc	tttgc	tttgc	6300
taatcaca	caagtaaa	tttgc	tttgc	tttgc	tttgc	6360
gtatcatt	tgtagct	tttgc	tttgc	tttgc	tttgc	6420
tgcagtgg	gaaaatg	tttgc	tttgc	tttgc	tttgc	6480
gccttctg	ggtttct	tttgc	tttgc	tttgc	tttgc	6540
agtatcata	acatagct	tttgc	tttgc	tttgc	tttgc	6600
attgc	ttcg	tttgc	tttgc	tttgc	tttgc	6660
ttgctt	taatgaaa	tttgc	tttgc	tttgc	tttgc	6720
ttatagg	gttta	tttgc	tttgc	tttgc	tttgc	6780
gtagccat	tatacca	tttgc	tttgc	tttgc	tttgc	6840
ggaattt	cttcaa	tttgc	tttgc	tttgc	tttgc	6900
ttatgc	caaggta	tttgc	tttgc	tttgc	tttgc	6960
aatctt	tac	tttgc	tttgc	tttgc	tttgc	7020
tcaccc	tgagg	tttgc	tttgc	tttgc	tttgc	7080
tcaagca	ttcc	tttgc	tttgc	tttgc	tttgc	7140

aacctggcata agttttgtat ttttagtaga gacagggtt cgccatgtt gccaggctgg 7200
cctcaaaactc ttgacctcg gtgatctagc ggccttggcc tcccaaagtg ctgggattat 7260
aggcaagagc caccgtgcc ggcttccaaa atatttaagc aatattatttgc cattttacaa 7320
ttttagtaat gcaaggaaacc aaaataaaac agaataatttgc agatagagaa gactttacac 7380
atcaatttgc aataaaacata taggaatgcc ctatatttgc aaatttcatg ggatgtacaa 7440
aatctacatt ttgattttgc tatatacgta tattttatttgc aatgatttgc cagagtataa 7500
cccaccaccc gcaactctaa kaaaatttgc gatgatttgc cgtcttggc agactgtact 7560
ttgatccatt tggcttaacc tggactata tggacttgc aagatacaga ctaatgaacg 7620
catctctagg tcccttgc ctccaaacaaa tacagtgc cacaatttgc aaatatttgc 7680
attctatttgc caattccctt tctaaatcaa caattttatgc catcatcaat tttataaata 7740
gccctgttc ttagacctt gatgatttgc atgtaatatttgc taaccttgc agtgcagact 7800
agttccaaaataa taatccata acgcccgtcc tccaaaggat cttggccctt gaccaatact 7860
tctgcctcct tctactgatt atgcccattt tctgctctca cagttttgc aatcttcat 7920
cttctcaaat gtctcacctt ccacaacttc ctacccctca ctcataggat gacctcactt 7980
cagttttat aggcaaaaaggc tgaagagttc tgatgagaat tccttgcattt catcatacta 8040
aaaaggcaga agttatctac ccatgttttgc ctcttccatg ttattgaaag ggatggctaa 8100
tttctctgtt tgactaagg atctcatgc aataggaatt ttatcctcgc aggagagtctt 8160
ctcttgagga acaaagcata ttttaatttgc ttgttgcattt gatggatccatg 8220
ggaagatacc taataaaatatttgc ttgttgcattt gttttcataa gatagggtcc 8280
attaacatta acaggtaggat cattttctac atttgttgc ttgttgcattt gttttcataa 8340
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gttttcataa gatagggtcc 8400
gaaaacaat attttctcca atttctacttgc ttgttgcattt gttttcataa gatagggtcc 8460
gactctctac aatcaatttgc atgtaagctt ttgttgcattt gttttcataa gatagggtcc 8520
aatggcttta aggtggaaatt ttggagtagt ttgttgcattt gttttcataa gatagggtcc 8580
gagaactgag agggagttggc ttgagttggc cataccatggc aatggatccatg 8640
cataccatggc aatatttgc atgagttggc ttgttgcattt gttttcataa gatagggtcc 8700
ccacacatggc cccaaataggc agaatttgc atctgttgc ttgttgcattt gttttcataa 8760
agatctctt tgagtggcata tctaaaaggc ttgttgcattt gttttcataa gatagggtcc 8820
tggtttgcct gaaatttgcattt gtaacttttgc ttgttgcattt gttttcataa gatagggtcc 8880
taatgtgtgc tctgctccctt ctattataat ttgttgcattt gttttcataa gatagggtcc 8940
ctgttagtaac cgcaggaaattt ggaacagaaaa acatgttgc ttgttgcattt gttttcataa 9000
acaggcagtg acattacaaa ttgttgcattt gtaacttttgc ttgttgcattt gttttcataa 9060
tatatcagat agaaatttgc ttctcttagtt ttgttgcattt gtaacttttgc ttgttgcattt 9120
aactatgggt cttaaatttgc aatgttgc ttgttgcattt gtaacttttgc ttgttgcattt 9180
tacaatatca gagctggagg tataataaag atcagacttgc ttgttgcattt gtaacttttgc 9240
gatgcaaaaacttgc ttgttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9300
cacgaagttt gaataaaattt accaaatgtt ggttgcattt gtaacttttgc ttgttgcattt 9360
ggaatgaata aacttatgtt gtttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9420
tttgcattt gtttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9480
gagacaatgtt tttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9540
tagataaaata gatgaatgtt gtttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9600
gagagaaatc ctcacatttgc ttgttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9660
caaaccacac catagtgaca gaatttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9720
agaagatcct gtttgggtt ttcccttgcga tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9780
gcatgatcct gctgatcagg accaatttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9840
gccacctctc ctttgcattt gtttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9900
atttcctctc agaacagaag accatcttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 9960
tcatcgccctt agtgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10020
tagccatttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10080
ttgttgcattt gtttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10140
ttcacttatttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10200
ttatcatgtt gtttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10260
gctttacttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10320
cgatcttcatttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10380
acctgacaat agtgcattt gtttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10440
cagagaagtc ttttgcattt gtttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10500
tgctgaaccc attgtatcttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10560
tgatttagggg aaaatccttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10620
gaattgttgcattt gtttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10680
ttatctaacttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10740
aagattgttgcattt gtttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10800
cagaaatttgc attgtatcttgc ttgttgcattt gtttgcattt tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10860
atgaaaataat tttttttttt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt gtaacttttgc ttgttgcattt 10920

ttacttactt	gggggggtggg	taaattttta	actaacttag	cctgagagaaa	attctgaaaa	10980
ttcatctcct	ttgaatatga	tggccacag	agccataatg	aggttcaaga	tagggtgat	11040
aattttacac	ttttaaaaaa	atgaaattat	ttctacttca	atatgggtc	ttgaacaata	11100
gctcagctac	ctttgtgctt	attgtaatca	atgtcaaaa	aatataaata	tatgggtcac	11160
tgatacgta	aatttgcgt	aagtgtccc	caaaccatc	ttgtaaaat	tttgagaagc	11220
ttctttttt	ttcatgtgat	aatactggtt	atcgctt	attctgaaaa	tcaactctca	11280
acaccccttt	atgagttcc	aagtgtatgg	atccaaattct	atcatcaaac	tactttaact	11340
gcattccagt	tatgtttcc	agtgcgtac	ctctgactt	ggactaattc	agttaggagtc	11400
aatagggtct	attaactgcc	ctgtgcatac	ccataaaatgt	cttgaacacag	tatacaaata	11460
aatagggtg	tataatgtt	tctggaggaa	atactcagtg	tttattttgc	ttcatgaagc	11520
tttctgtg	ccattat	taaaataagg	gtgttctagt	agttctgaa	gctgctata	11580
agaaaaatcaa	tctcagagag	gctacctaag	gagacacatt	aatccatgca	ttattagtca	11640
accagatgac	tgcattatga	tacatatgt	tttagcccta	ataaaattgcc	tcataatgtt	11700
gcatat	tat	tctgatatta	actaacacca	tacagaat	gctacagaaa	11760
atgtctgtat	gatatcaact	actatatcta	ctataaagt	ttcaaatgg	tttccagaat	11820
attaagcaga	tgacttgcag	ttttgtaaa	atggggcat	attctgagt	ctcatttgc	11880
taagtctgt	tctagccaa	tcctcttca	gaaacagtg	ctgaaaaaga	aggaaggcaa	11940
atttcggtg	caagaat	acaattacat	tttagcact	tttggtttt	tat	12000
gttttggcaa	tgatgggg	acaacagagg	aagg	gcagagacaa	caatgaagca	12060
gaaaagctgg	gaatttgat	acacagagtc	ctcactt	tttgagg	cagcatcatt	12120
tcatcggtgg	ggtatttctt	tat	ttat	tctattag	aaagcattgt	12180
attccattt	tgccaaggcag	aatgg	tat	tgaaacc	tttgaaaatt	12240
gat	ttttaga	atgatgca	aaaatact	at	taaataact	12300
tttattttct	agttatgaga	attgct	cttgcag	ccat	tttgcacat	12360
ccttttagt	aaaatgact	ac	tttgcatt	aatt	taataaaa	12420
ttctat	tttgcattt	tttgcatt	catgt	tttgc	tttgcattt	12480
ttcttggag	ataggataga	acataa	atcg	acat	gaataatt	12540
gtcacagt	gtgtctgatt	tttag	aaac	ttt	tctcataaa	12600
agagaattat	tat	tac	aaat	ttt	tgaaataat	12660
tgcccaaa	gtt	ctc	at	ttt	tcttgc	12720
tctcttctt	att	ct	at	ttt	tttgc	12780
tggtggaa	ttt	tt	tt	ttt	tttgc	12840
ttatgtgata	ttt	tt	tt	ttt	tttgc	12900
gttaacataa	ttt	tt	tt	ttt	tttgc	12960
aatttgcata	ttt	tt	tt	ttt	tttgc	13020
cagtcgcaaa	ttt	tt	tt	ttt	tttgc	13080
agtaaaattt	atgagg	tgt	cct	ttt	tttgc	13140
catttctt	caataa	at	tg	ttt	tttgc	13200
aatcattt	catata	ttt	ga	ttt	tttgc	13260
tatata	attt	ttt	ga	ttt	tttgc	13320
atgacctt	aa	ttt	ga	ttt	tttgc	13380
atacataa	at	ttt	ca	ttt	tttgc	13440
agcattt	act	ttt	ttt	ttt	tttgc	13500
atgtat	ttt	ttt	ttt	ttt	tttgc	13560
atgttccc	ttt	ttt	ttt	ttt	tttgc	13620
atgttccc	ttt	ttt	ttt	ttt	tttgc	13680
aat	ttt	ttt	ttt	ttt	tttgc	13740
ttat	ttt	ttt	ttt	ttt	tttgc	13800
cagaataaaa	att	ttt	ttt	ttt	tttgc	13860
ctcttcat	ttt	ttt	ttt	ttt	tttgc	13920
aatgtgac	ttt	ttt	ttt	ttt	tttgc	13980
aagcactt	ttt	ttt	ttt	ttt	tttgc	14040
cattttgt	ttt	ttt	ttt	ttt	tttgc	14100
tcctcccact	ttt	ttt	ttt	ttt	tttgc	14160
ttataccata	ttt	ttt	ttt	ttt	tttgc	14220
tatattttaa	act	ttt	ttt	ttt	tttgc	14280
gaaaaatgt	ttt	ttt	ttt	ttt	tttgc	14340
agatatccag	ttt	ttt	ttt	ttt	tttgc	14400
tgcataactt	ttt	ttt	ttt	ttt	tttgc	14460
ctgtt	ttt	ttt	ttt	ttt	tttgc	14520
ctgcttatt	ttt	ttt	ttt	ttt	tttgc	14580
ggagaaaaca	ctt	ttt	ttt	ttt	tttgc	14640
catttattt	tat	ttt	ttt	ttt	tttgc	14700

taaatacaga	ttctttctc	taactatgct	gacttttgc	caacaactcc	acccaaaaga	14760
gttgcacatgc	ccacatatat	tcttctgcta	ggaatcattt	caaacactga	aaaaaccccta	14820
ttaatttagaa	ttaaaaataaa	tattaaacat	ttaaataatc	agagaaaaat	tactctgtgc	14880
ttcctaagaa	tggaaatact	tcacagtaat	actgttgctt	taatttgaca	ataaaaactat	14940
ttaaaaataa	gaccaccct	acagtctggc	aatgtatTT	tcacaaataa	gacaaagtgg	15000
cacaattact	gaaatatgtt	atgttaat	aaatcattt	aagcactaaa	acatgaccat	15060
ataaaataat	gttgaatact	atacgtagaa	aattgattaa	aataaaaaat	gctaaaatta	15120
acgcagaaaa	tagtggcagc	taaaaacatg	aagtacaata	ctgttatgt	gtctatattt	15180
aagacagctg	aaaaagtcta	gcacgttcat	acattgggt	attaatggta	taaaaatttt	15240
ctagaaataa	atacaatatt	attaatattt	taaggagtca	aatgttcta	taatTTTatt	15300
cagctattcc	aaatttaaga	atctactgaa	ggcataatta	gaaacttgt	tgaagagtt	15360
catgggtttc	atTTTataat	caagaaaaag	caaacaaatg	atgttcgaaa	tacaagttagc	15420
tcccacttag	caccgtacca	tgtaacagaa	acctttctat	atggcaaacg	tgtctgttt	15480
tgcatggttc	ttacttagat	attacaaaat	ctcaaagtga	ggactgcctg	cttggataa	15540
ctggtaaagc	acaggcatct	accagatggg	gtagaaggt	tttctaaata	actagaatgt	15600
catcagccctt	gatcatgaag	gaaagtgtaa	agtagtctt	aattttttt	tcttagaaatg	15660
ttatacagct	attaaagtga	ggatgttctt	agtaaagaaa	aaggtaaatac	atTTTgtta	15720
atataattatg	tgcaagagaca	tatccatgc	tacatatgg	tatcttgtaa	aatatcagg	15780
ataaaagagaa	acaaaactta	ttaatagcta	tgtccaaatt	ttccatgata	agcaattgtt	15840
gctattacat	tgagaaaata	agctaacaaa	aactttgtat	gcccgatTT	attcagatct	15900
caggatata	aattacagga	tgcttattt	taatagtaga	atTTTaaaat	aaaagtagga	15960
aaaagcaagg	agttaaaaac	ttgtttaga	gcgtgatggc	agaagaggag	gtgctcagg	16020
ctggTTTCCA	gaacagaatg	accagttagc	aacttctcac	agataagaat	gcctaggtaa	16080
aaattccata	acccaaagggt	gagggtgggc	acactcttgg	agcacagaaa	ctgggaaaag	16140
tcatattaga	agtgtaaagg	gagcagttc	acttgactg	tgttgcTTt	ttcccacca	16200
aacagtgtca	cactgaaagg	gatttctgg	gcctgcgtt	tctagtggag	aaaagagagc	16260
ccacggcaga	catccaactt	ccctgtttc	cagggcactt	cccaaggggc	ctagttctgt	16320
ctaattttgt	ggggaaataat	ggaggaattt	gcagggctt	accctttag	gtcagttcat	16380
aactacccc	ttctacaacc	cattctgtat	tccctttacc	ttcagcaagc	accttagcag	16440
gacacggtt	tttacctgg	ggagtgcac	aaatcttcat	tcctgatggg	tctgggcat	16500
ttgttagttt	gcctggattt	gtttgttga	gtttctact	gaccttaatc	acaggacatg	16560
gtaatactgt	gacgtcctat	gggatcttctt	gtattccaca	aatactctt	cttaacacccc	16620
attgtggagt	agtagtttga	tttcatctt	atagtctgg	tcagtcaccc	cagccaaacac	16680
tgtacctccc	ttcttagcct	gttgacttag	aggcaggaag	agctcaaagt	ggccagggtt	16740
caaacttatac	ttccagttt	atggaatcat	ctttctgtct	tctggtggc	aagttcctcc	16800
ctccggaaacg	aagacctcta	ggcccacaga	acataatgtc	acagaaacag	gaaggaaaat	16860
tttgcTAGt	ggTCAGTGA	ggtgatgtt	aggcctata	tgcagtgtc	cgatcttgg	16920
ttcctgcatt	cttgatcct	ggctatggaa	gaaacagttac	cacatattgg	atgcctactc	16980
agagcataca	aagacttctg	gagaaattt	ccccaggcct	gcattttttt	ttttttttt	17040
tgtaaaatga	agtctctgt	tgttccag	aggcctata	tgcagtgtc	cgatcttgg	17100
tcactgcaac	ctctgcctcc	caggttcat	cgattcttt	gtctcagcct	cccgaatagc	17160
tgggattaca	ggcatgtgccc	accacggctg	gctaactttt	gtatTTT	tatggatgag	17220
ggttcaccaat	gttggctagg	ctggtctcga	actctgtacc	tcaagtgtac	cacccgcttc	17280
agcctcccaa	agtgtggaa	ttcaggcat	gagccaccgt	gcccgcccc	tgtataaaag	17340
tatcacctag	ttgtttaga	tgTTTacttca	aaaggctttt	ccaccttct	atcaatccat	17400
ctgcttttag	ataatggaa	acatggtaag	acaagtata	cccatgagca	ttagccccact	17460
gtcttatttt	ttagctgtaa	aatgagtgc	ttggtcagag	gcaatggtct	gtggatatacc	17520
atgacagttg	ataaggcatt	ctgggagttc	acgaatggta	atcttgcag	aagcatttgc	17580
tgagagatag	gtaaacctat	atccagagga	agtgtcttaat	gtgactaatac	cactcttagaa	17640
ttctaattcac	ccaaagctt	tggattcctc	cctctacatt	aaaccaggga	agatctggca	17700
tttcaaggc	gctcacagta	ggccatcctt	tgatccat	ttcagctaat	aaactattag	17760
aacattttt	aactccctga	tttgcataat	tgagtgcaga	atctctgtt	agtggggccca	17820
tatcaataca	ttcagcctga	tccaaactt	tattcctccc	accattatcc	cacaccctta	17880
atatccatc	ctacacctgt	tctccagaat	tctgcttata	taaactagaa	aactcaagca	17940
gctcttttag	ggttagc	acctcctt	agactttgg	cttacaccaa	ctagcaccag	18000
atcttggaa	gagaaaagg	attaaagtct	gaatgtattt	cctaaagaag	aagaccaacc	18060
aaaatagtt	atTTTataat	ttggctattt	taagtccaaa	gctactctt	aaaaggaagc	18120
catatttaac	cactccatgt	aatatgtat	tatgtcttga	catacagaac	taggtatgc	18180
aataattaaa	ggtattttgg	caccataatt	ttttctaaa	aacggagaaa	tgtgcaaaatt	18240
atTTTACCTAC	cagtttacac	aaacaggca	agcaaaaaaca	aaagcactt	tcaagggtt	18300
caaaaatgt	gctacagtt	aaaatctgca	tatactctt	actttgtcat	tcttaaaaat	18360
gtgtttctgg	gtagttgg	tactgaaata	ggtgaaagct	gtactgtat	gtcatttttta	18420
aattttgtaa	cactttt	aacattttgt	aacattttgt	gaagaaatta	tgtttcttca	18480

cttaaggaga taaaattgtgt	tcattataga actttgaaag tgaacgacca aaccaaatgt	18540
gcaaaaatgt actgaatatgt	aattacttta cacttatgtg tgaaaataat cttcaacatt	18600
atgtattcca tacacatgtt	tttgggtgtt taaattatct ccgggcttc ctaaattcaa	18660
gttggagaag cagaattagt	tagatttata gcacactcac tataaaaaac atatttgaa	18720
aacttgtttt accacatctc	tgttcagctg gatggggctc ctgtgaaaga acaaagcaat	18780
tacatagaat tttacacaga	attgtaaaac ctgcaaattt ttttaacat tttctttacc	18840
ttatctgata gcacagcaca	tacgttgcaa ctcttactct gaaaactact atcacttctg	18900
aatgtttattt aaataacagc	cataggtgtg cgaaggattt acaaatggat agccacattt	18960
actgccaatg aaaaacaaag	tttatttaac atgtcagta tctttgtat atatataatta	19020
ctccttattt ggtacagatg	aaaacgtga agtcagaag aatgaataa tgtttccag	19080
gaacacacgca agagcttagt	ttcaattaaa atattattct cgtactact tctgggttga	19140
ctctagaaac accccaagcc	acatgggtt gttaagatgt ttaactgggt aattatttt	19200
atgggtaattt atgttcaga	cagtttaca attttacaga tataaaaata tatatgaaa	19260
agatataatat atgttgcata	gcatatataat atattcatat ttcatatgaa gaaggcaca	19320
gaggaacact cagtgcctag	gtaaaaagaa tggagttgga tagtagtta atagtggtg	19380
atgagataag aagtagtcat	ggtcatgact tactttctg gtgtcatgca atttctattt	19440
acagcatttc ttcaagttt	taattttat tgagaataaa ttaataaaac tctgttaagat	19500
tttaagataa aagatctccc	aatattttca aagcatcatt acaaccttct taaagttt	19560
tagcaatttg atcaatttct	ttattcttat tttcattgtt gaaaaaaattt gattttttag	19620
gtatyttaaa agtcatagag	aaaatatttg gccatggaaa gggtccaaa tgggaggagg	19680
aaattaaaaaa ggactgagat	ttctgaaaca catacattgt tccaaggag taccaagcat	19740
gctacaaagg aggcaagagga	ttgaatgact gaagaaaaga ttttgaagga cacttgagg	19800
tgttcttctt acacccatca	tcatcctatc tcaggagggt gaatggattt gagagattat	19860
aaaaagaggattt attgtgatgg	ccttgatgtt agtccaaac atcctttct agtctatgt	19920
attgataagg atgctgaagc	tgttcatggc tggaaagg gaaggggaga ggagagttat	19980
cccattgtg tactgtatct	gagtgaettc acctcagagg aaaggAACAG agaagcatct	20040
ctaaaccaca ttctcaaattc	caaatttacc agcaataatgt ctagtaattt tcatatctga	20100
ctccccatgtg tagactctaa	ttgagtcagg atccaaaagg agggcaattc aatggctcct	20160
ttaagttcta acatgtatgaa	tgtagaagta ttccaaattt ttagctaaaa taatatttat	20220
cataatagttt atttactgt	ctactaactc actgcaagac ttatgaaata actactat	20280
gatgctctgg ccaggatctg	agatgtccac aattctactc atatgtttt gctaaagaga	20340
atttaatgaa aagacagaga	ataaaaaaggtaa agttaaag gaatcagcca aaggggttaa	20400
aaaagtcaag gctaactgt	gaaaaccatt tccatcttta ggtctgaagg accaaggaca	20460
gtggcttcgtc tCACAGAGCT	cagttagagc tgaagccagg aagaaatccc ttgccaacat	20520
cttcccaaga cacacatccc	aagacattct ccaggagggt atgcagccaa ttagcacac	20580
ttrtaatgaa gcaggaaagg	aatgagaaaag aaatttgggtt aattccctat ctttcccact	20640
tctgacatcc tagagagtct	tctcaaaagc caaggtcagc tggaaaccag acaccatagg	20700
tgccctgtt accaggctca	taggagccca actccagggt aacaatgcaa ggcacagaag	20760
agaagacaat agatgtgaat	gggagtaggg tggaaaaggga gaatcaccag cacacctata	20820
ttatctataa atccaaaatgt	ttccacactt acacattact atcccttataaaaagataaa	20880
tgtacaaacat tcaaaattgt	tatatcaact gttcaagttc aacaaaaggat agtttatctt	20940
ccaatgggtt tccagactct	gaaaccttct cccgataccca tccactatac ttgctgttaag	21000
caggaactt atatttccca	aagtcttaaca tattaaattt tttgggagta aagtaaaaata	21060
aataatgaaat cccattctt	catcactt gggggatgcg attccctgaa gacagtgtgg	21120
agttggcaaa actcatgtgc	tccgtgagct taggattttc agaagaatcg aatatttgc	21180
gttggtcata atttgcata	cttttgcata gaaataattt acctgcaggg atttggtctc	21240
cacacagtaa ttgcattata	aactctacag gttctcaata cccattctgtt aatcatgaaa	21300
tcatgcttctt ttctcaggta	ggatttagata tatttttaaa aaagttttaa atccggcagt	21360
aaatatttttta attgatagaa	ataaaatggag ctgaaacata ttctgatgaa ttagaaaaatg	21420
tttatattaa agtggtgta	ttcttttcag agtactgggt tttacagagt gtttgcataa	21480
tgaaaagaaa atagggccaaa	agaatttttta ttcatcaat aatttttgc tgcctccaaat	21540
tcaggagggca caaatctagg	tgcgtctgag tctataacta gatcagataa aatcccttga	21600
ccctaaagga gctcatgctt	tgatgttggag ttctggttcta gactgacact ggatcattca	21660
ccaaataataa ctatgtacct	caatttgcata agcatgtatg acatgacaga tatattttaa	21720
aaacttttac tgaatataact	ttctcatcaa gagagaatgg ctattaatca aattttatgg	21780
ataagaaaac tgaggcacac	agaaatttaat aacttgccttcc agggccacaca ggttaggaact	21840
agaggcactg ggtttttcac	ccaggaagtg tgactgcaaa ctcccagtaa ctgtcccact	21900
ctgctggctt tcaatagtaa	actatgcata ctgtttatac ttgatttgcata acaatcaaga	21960
tctggtattt aatcatccctt	ctgattcatt ttgtgtttaac aatgcacactc ttacttattt	22020
taagagctgt atcatgtat	attttatacg cttttcataa agaatgggtt acttttaggt	22080
ttatagccca tattacttaa	aaataataaa aatatgtttt gtgaagctgc cttctctaga	22140
tttcttttcc cactccaaat	tgaaccatga ccactttcca aggagacacc actcttgcata	22200
aagacaaaga ccaaccttct	agactcttgc taggacagaa tagaagaagc aagcattcac	22260

atggaaatgt	ctcagtgtag	ggtcgtatgtc	aggagagcac	aggctcagat	gcctccaaag	22320
ccatttctgt	ccaattgcag	aagcttctcc	ctgagatttc	cttgc当地	gcagtaggat	22380
attgttct	tgtttttca	gaagaaagac	tatttccaca	aaattctcag	ttaaatttatt	22440
tggactattt	attgttgca	ttacttcaac	ttccaataga	aaaaaaaaaa	gactcataaa	22500
aaaacagatg	aatttggaaag	ttatttgaa	gggtgattaa	tatttcaata	aaaaagccca	22560
ttaaaaatgt	aatttaggcgg	ggcacggtgg	ctccccgcctg	ttatcccagt	actttgggag	22620
gctgaagcgg	gtggattgcc	tgaggacagg	agttcaagac	cagcttggcc	aacatagtg	22680
aaccccgtt	ctactaaaaa	tacaaaaaaa	ttagctgaac	gtgggtgtgg	gcacctgtaa	22740
tcccagctac	tcgggagggtt	gaggcagcag	tattgctga	acctgggaga	cagaggttgc	22800
agttagctga	gattgcacca	ttgcactcca	gcctgggoga	caagagtgaa	actccgcccc	22860
acacacacac	acaatgaaa	tccagcacca	aaacagtatt	tgcctatttgc	caattaaat	22920
gtaccatcaa	acagagggat	agagaatgtg	gtatattata	tatatatcat	acattataat	22980
gtgatataat	atataataca	ccacagtttc	tttatccact	tgttgtttga	tggatataata	23040
catatatattt	atatgtataa	tatatacataa	atgtatata	tgaatacta	ctcagccata	23100
aaaagggatg	aagtatggc	attcacagca	acttggatgg	gattggagac	taatattcta	23160
agttagtaa	ttcagaaata	aaatacaaa	tatagtgtgt	tctcactcat	aagtggaaac	23220
taagctgtca	ggatgcaaaag	gcataagaat	gacacaaagg	acttcgggga	cttgggggaa	23280
acggtgagg	aggtagggga	taaaaggcta	cacattgggt	tcattgtata	ctgtttggat	23340
gatgggtgca	ccaaaatctc	acaaatcacc	actaaagaac	ttactcatgt	aaccaaatac	23400
cacctttcc	cccaaaacct	atggaaataa	aaaataaaaaa	taaaaaccaa	aagaaattca	23460
attggaccc	gtgagctta	acaaggtaat	tatgattcta	ttgatttagg	tgacttatct	23520
tctaacttat	tacccctaggc	agaagccaa	tgtccttttgc	tgttagatga	gataatacgg	23580
tggacaagg	ttgaacatttgc	aatttacaga	gtgctttatt	tatgaatgt	cctgtttcca	23640
gtgttggcat	ggtagtttatt	aaaaaaaaatgt	tagtataaca	ttgtcaagtc	tggatatgt	23700
tcatagaaga	aacacgtcga	ttgttccatc	atgatctca	tggtcaactt	tatctgccc	23760
aaatgtttag	agattcagg	caggatattt	taagtatgg	ctcagtcatg	ttgttaat	23820
gattatgatt	tcactttgtat	ctgttacatg	ttttaacaaa	ttcatattga	acttgggttc	23880
caatgtttt	tcttagttt	tcaagtgtt	gaatgaggta	gagttaaaac	aacagcattt	23940
tagtttgg	ataagtttct	tcatcatttt	atccttatttgc	ccccccccc	caggagccct	24000
ttgatacact	attgtttct	ttatatccct	atcaacattt	aataccatca	atacagggtc	24060
acttggat	ctatataccaa	tgcaaatgt	tatggtttat	gtattatgt	cataatgtatt	24120
ataaaatcaa	atatgtataa	tctacttca	ttcccccaaga	accaattttt	ccccccctgg	24180
acaaaaat	acaaaatgaa	gaaatgttagc	ttggtcttta	aagtaaaaaa	agattttgt	24240
ataaggggc	cagctgagg	agaataaagt	gaagaaaatgt	ggcacacatt	agaggggtgag	24300
ggaagatctg	gaaaatgtc	tccagttcac	agggggcaggc	aaagcgattt	ttgttctaca	24360
gggatgtata	ccagccatca	gtcccac	cctaaggcag	ttcagttgt	ataaacac	24420
agggtttcaa	aggctgatac	tttagatccc	acattcaaaag	gtgtgttgg	aaacaaagaa	24480
ttacagttc	aaagaaaaagc	aatgttaca	accatgggtt	caagaaaaatgt	ctaaatgtac	24540
acatataaca	aagacttgc	aaaagataaa	agataaggct	ctttaactat	caaaagactt	24600
gcagaaaaa	accacagaaa	accattttaa	atattattgc	ctttagtata	taaaaaaactc	24660
tatatttagt	tagatgttta	aagcatcaat	cacatgctca	ctaggctatt	tcttaatgtc	24720
acatgtattt	acattttgag	agaagagggaa	gaaatagcag	atgacaccac	ttgggttaat	24780
cataaatgac	aaacctaataat	gcattttat	ttccttttat	ttagatgt	tttgaagcc	24840
agcaaaacaca	atgttaaaga	aaaaccatac	agccgtgact	gagttttgtt	tcctgggact	24900
gacagatctt	gttgagctgc	agtcccttct	ttttgtggta	tttcttagt	tctac	24960
cacagtaat	ggcaatgt	gcatgatctt	gttaatcaga	agtgtact	caactacac	25020
tccaaatgt	ttcttcctca	gtcacccctc	ctttgttagat	ctctgttata	ccaccaatgt	25080
tactcctcag	atgttggtt	actttttatc	caagagaaaa	accatttctt	tcatcggt	25140
ctttatccaa	tttcaactttt	tcattgtact	ggtgattaca	gattattata	tgtcacat	25200
gatggcttat	gaccgctaca	tggccatctg	caagcccttg	ttatatggaa	gaaaaatgt	25260
cagggtgt	tgcctctgtc	tggctgtgtc	tccttatatt	tatggcttgc	caaatggct	25320
aagcacagac	caccctgtat	cttcgtgt	cctctgtgt	acccaaatgt	atcaaccact	25380
tttactgtc	ggacccaccc	cttttagtcc	tcgcctgtc	agataactt	gtcaaagaga	25440
ccgcatgtt	gggtgtggct	gggtccaaacc	tcatttgc	tctcaccgtc	atccctcat	25500
cctacactt	catcttact	gccatttgc	gtatccacac	tgctgagg	aggcgcaagg	25560
ccttctccac	ctgcgggtct	catgtgacc	ctgtcactgt	cttctatgg	acactgtt	25620
gcatgtac	gaggccccct	tctgagacat	ctatacaaca	ggggaaaatt	gtagctgtt	25680
tttatatctt	tgtgagtccg	atgttaaacc	cattgtatca	caggcctgagg	aataaagacg	25740
ttaaaaagaag	tataaggaaa	gttattcaaa	agaaaactgtt	tgctaaatgt	ggtagatatt	25800
ttggtcatag	gcgttggaaat	ctgttctt	tatctgacca	attaatgtac	attnaaaatt	25860
aacaaatcaa	tctgtcatttgc	agtgtttttt	gtcctttgt	atttgcata	gggacttaaa	25920
agtgtatgtc	aaatttattag	ctagagctt	cactgccc	tcaatgtaaatt	gaaaatgtaa	25980
gcatagaaat	tcaaatataa	gactagaaga	cattgttctt	gctctgtaaa	aagtaatgtaa	26040

gtacagatga ttccattatt	agtactcata	gtactaatat	atcatagtc	tatatcatag	26100
tactaatata tcatagtact	atatcatagt	actaatataat	catagta	tatcatatgt	26160
ctaatataatc atatgtactat	atcatagtc	taatataatca	tagtactata	tcatagtact	26220
aatatatacat agtactataat	catagta	atataatcata	gtactatatac	atatgtactaa	26280
tatatcatag tactatataca	tagtactgt	tcatagtact	aatatatacat	agtactgtat	26340
catagta	atataatcata	gtactgtatc	atagta	tatatcatag	26400
tagtactata tatacatatgt	ctataatca	tagtactata	tatacatatgt	ctataatataca	26460
tagtactata tatacatatca	aaatatacata	aaagacagaa	ttatcaattt	ctgagaaaaat	26520
gacagctat tacattttt	taaatgttt	tgttgtccc	tttgaatgaa	tggaggat	26580
ctacctaact tctatgtatt	cagaatctca	catgaggcag	cacatttagc	acacttgtt	26640
agttgtatc ttctatgctt	atcttcgtga	gtctgtttga	aaagttggat	aatatggcaat	26700
gagtgactac atttgctaat	tacatttat	aatacttaat	aagccctcaa	gctatttcc	26760
agaaaacttgc caaccatagt	agatgagggtt	tctgttttc	ttttatacat	attatataaga	26820
aatgaagatc ctttaggtgt	actttggatt	attactttaa	gttataatcct	tagaatgata	26880
tcatagatcc aatgaaaaccc	tatttcataa	ggctttcat	ggaaatgcca	aatttcctt	26940
agaaaagagtt ttaattgtt	tgtatgttgt	accagtaagg	atcccatca	acatgtgaac	27000
tcaattcttc actcttttat	taccattgcc	tgttgtcc	gaataaataa	attgaattga	27060
atgttttatac cacgaatgca	tgtggatatt	tgtgtgttt	atgtgtgtat	gtaggtgtgt	27120
gtgtgtattt tgattacaat	ataacatttt	cttttttta	attgcttatt	caaattttt	27180
tggtgagaat caaggtcgta	tgcctttttt	ttttttttt	tttttttga	cagagtattt	27240
ctctgttgct caggctggag	tacagtggcg	cgatctcgcc	tcactgaaag	ctccgcctcc	27300
cgggttcaag ccattctcct	gcctcagcct	cccgagtagc	tgggactaca	ggcgccccgc	27360
accacgcccc gctaattttt	tgtattttta	gtagagacgg	gttttccaccg	tgttagccag	27420
gatggcttgc atctcctgac	ctcggtatcc	gccgcctcg	gcctccaaa	gtactggat	27480
tacaggcatg agccaccgcg	cccagctgct	tcacctgttt	ttaaattgtg	gtgtttat	27540
gttttcttat ttatattcaga	tatttcataa	tgtatataaac	atgttttgc	tatataataca	27600
ggtgttagcaa ttgtatataat	atacacatga	attccatttt	atctcttcta	ttgtgtctt	27660
tccaaagtta atattttatac	attttaaattt	ttctcaatata	ttttttattt	cactttattt	27720
tgttttttt tacttttagg	aagtttaagtt	tatgccacct	atgatcatgt	ggatttttac	27780
atataattttt aaattttttt	catggtttcg	tttgttaattt	ttttttat	aactattttt	27840
aatatacatc taaaatagaa	aataatataa	agactcctt	gtaagcacaa	ctaataacta	27900
ttaaatcttc acagaatgcc	ccattttttt	cagttattt	agaaattttaa	aacttacaaa	27960
gatacatctc tcctgactga	ttctgtctct	tgtcccattt	cttgctgtaa	actctattac	28020
aaatttaatg ttcttattcc	ttctgttagag	ttacatcatt	ggataagggt	ttatataattc	28080
atagaaatat gttagtactga	tttgcatacat	ttaaaacttg	aaataatgc	atccactcta	28140
tgtatcttc tgcaattttat	ttatcttgca	ctaaatattt	ttggatgtat	ctgtggtgaa	28200
tcatgcagca gcaattttta	gtggatagtt	ctatgccaat	atataaataat	gcctcaat	28260
gtgttatttt tctacgattt	ttttttttt	ttgagacaga	gtctcaact	tgttgccca	28320
gctgggggtgc aatggcacga	tctcggtca	ccgcaaaactc	cgccccccag	ggtaagcga	28380
ttctcctgtc tagcctccta	gtagctggga	ttacaggtgc	ccaccacat	gccaggctaa	28440
ttttgtattt ttaatagaga	cgggatttca	ccatgttagc	caggatggc	ttgatcacct	28500
gacccttgtga tcctccgc	tcagcctccc	aaattgtattt	ttatttctac	tgttggtgaa	28560
gagttacatt ttccagttt	ttgataatata	gaaaatccat	taataaaatgt	tttggatgt	28620
attatgaatg tgcattacatgt	aaggaggatcc	tctcaggttag	gttggatatt	tcttggtcat	28680
aagatataag tatctttaaa	tttactatata	attgccaatt	gttttactga	tttatatttc	28740
tatcaacagt ttgtgaggt	tacctttaca	tcaactttat	tttgccaaaa	ctagggtgtt	28800
tcagacttta aatgttttct	aatctgtatgg	gtaaacaat	aatttcatt	ttgtgtttaa	28860
gtttaatttgc ttatttataag	tgaaactaag	tatttttaag	tatattgtat	ctaatactcat	28920
attttctctt cacttatata	tccatatact	ttgttaactt	ttcaggttag	gtgattttgc	28980
tttatttcat acatttataa	tagtcccta	tgtaccaaata	gctacatatac	aattttttg	29040
tgttatatagc cattaacat	taaaaaccta	ggtaataat	gccaggcagt	gtggctcaca	29100
catgtatacc cagtactttt	ggaggcgtag	gcaggaggat	tgcttgagcc	caggagttt	29160
aggccaaact acacaagata	gagactctgt	ctttacaaaa	aaataaaaaaa	aaatttagcca	29220
gttgtgggtt gatgccctt	ttgtccca	tacatggaa	gctgaggcag	gaggatca	29280
tgagcctggg aggttagaggc	ttcagtgaac	catgtgtca	ccacgacact	ccagtctgga	29340
tgacagagga tggatgacag	agcctggat	actccagcc	ggatgaccct	gtctcaat	29400
aacaaaacaa accccaaaaaa	accttaggggt	agatgaacag	atgaacgtt	tttcttgaat	29460
gcctcagccaa atagccatag	gtatctattt	ttctggtagc	tgcaatttga	ttgggacaag	29520
tttggaaattt cagctcattt	gacaagtattt	ttcataatag	atattttccc	agttaggattt	29580
gattttttt ttccacatag	tgaatgctt	tatgtgagca	taaaatgtttt	cccacagatt	29640
gggatttggat aaattatata	aagtgtatc	ctagtatatt	gcttgc	tgaaataactg	29700
gttatttaccc aaatagaaaa	ttaacat	aaagagccaa	caggctgggc	aaagtggctt	29760
aggcctgtaa tcccagcctt	ttgggaggcc	aaggcccaca	gatcacttga	gttcaggagt	29820

tcgagactaa	cctgggcaac	atgagggaaac	cctgtctcta	aaaaaaccac	aaaaattagt	29880
taggtcttgt	ggcaagtgtc	tgttagtccaa	gctactcagg	aggctgatgt	gggaggatgg	29940
cttcagcccg	ggaggccaaag	gttgcagtga	gctgagatca	tgccagtgca	ctccagcccg	30000
gttaacagag	taagatcccg	tcttaaaaaa	ataataaaaat	aggccaggtg	cagtagttca	30060
tgcctgtaat	cccagcactt	tgggaggccg	aggcaggtga	atcacaaggt	caggagttcg	30120
agaccagcct	ggccaagatg	gcgaaacccc	gtctctacta	aaaatacaaa	aattagccgg	30180
gtgcagtgtt	gggtgcctgt	aattccagct	acttgggagg	ctgaggcagg	agaatcgctt	30240
gaacccgggg	ggcagaggtt	gccatgagcc	aagatcgtgc	cactgtactc	tagccttgg	30300
gacagagcca	gactccaact	caaaaataaa	taaataaaata	aataaaataa	taaaatataa	30360
tataataaga	atgccaacag	aataattctg	atttcagcta	cagtttatgc	ttttttatga	30420
gaaattcaac	tccaatctt	taattttatt	ctccttatt	ctgttgttca	gacattgcat	30480
attaaaaagg	cattttttt	cttaaccaat	agttttatcc	ttattgtgt	ccttattctg	30540
gagagaactt	catatttcaa	ggcacagccc	aaatatcaac	tccctgtcc	tttcatgagc	30600
aaacagttgc	acataacccat	tcctgtgatt	gttagatata	tgcaactcta	tttgtctgtt	30660
agtcacacag	aaaattcata	ataatacaat	gccatgtttc	ctaataatatt	ctgagagttc	30720
tgtgagaaat	gaaacctatt	gtatcggtt	agtactgaag	atggccgaat	aggaacagct	30780
ccagtctaca	gctcccagcg	tgagcgcacgc	agaagatggg	tgatttctgc	atttccatct	30840
gaggtacccg	gttcatctca	ctagggagtg	gcagacagtg	ggccgaggtc	agtgggtgcg	30900
cgcacccctgc	gcgagccgaa	gcagggtgag	gcattgcctc	actctggaaag	cacaagggt	30960
cagggagttc	ccttccttag	tcaaagaaaag	gggtgacaga	tggcacctgg	aaaatcggt	31020
cactcccacc	ccaataactgt	gctttccaa	cgggcttaaa	aatggcaca	ccaggagatt	31080
atatcccgc	cctggctcag	agagtcctat	gcccacggag	tctcactgat	tgctagcaca	31140
gcagtctgag	atcaaactgc	aaggcggcag	cgaggctggg	ggaggggtgc	ccaccattgc	31200
ccaggcatgc	ttagttaaac	aaagcagcca	ggaagctcga	actgggtgga	gcccaccaca	31260
gctcaaggag	gcctgcctgc	ctctgttaggc	tccacctctg	ggggcaggac	acagacaaac	31320
aaaagacagc	agtaacctct	gcggacttaa	atgtccctga	cagcttggaa	gagagcagt	31380
gttctccag	cacgcagctg	gagatctgag	aaagggcaga	cttcctcctc	aagtgggtcc	31440
ctgacccctg	accocccgagc	agcataactg	ggaggcacc	cccagcatgg	gcagactgac	31500
acctcacacg	gccgggtact	ccaacagacc	tgtagctaag	gttcctgtct	gttagaagg	31560
aaactaacaa	acagaaaagga	catccacacc	gaaaacccat	ctgtacatca	ccatcatcaa	31620
agacaaaag	cagataaaac	cacaaagata	ggggaaaaac	agagcagaaa	aactggaaac	31680
tctaaaaaagc	agagcacctc	tcctcctcca	aaggaaagca	gttcctcacc	agcaacggaa	31740
caaagctgga	tggagaacga	ctttgacgag	atgagagaag	aaggcttcag	atgatcaa	31800
tactccaagc	tacgggaggt	cattcaaacc	aaaggcaaaag	aagttaaaaa	ctttgaaaaa	31860
aatttagaag	aatgtataac	tagaataaca	aatagagaga	agtcttaaa	ggagctgatg	31920
gagctgaaaa	ccaaggctcg	agaactacgt	gaagaatgaa	gaagcctcag	gagccgatgc	31980
gatcaactgg	aagaaagggt	atcagcaatg	gaagatgaat	tgaatgaaat	gaagcaagaa	32040
ggaagttt	gagaaaaaaag	aataaaaaaa	aatgagcaaa	gcctccaaga	aatatggac	32100
tatgtgaaaa	gaccaaatact	atgtctgatt	ggtgtacctg	aaagtgcac	ggagaatgga	32160
accaagttgg	aaaacactct	gcaggatatt	atccaaagaga	acttcccaa	tctagcaagg	32220
caggccaaca	tccagattca	ggaaaatacag	agcacaccac	aaagataactc	ctcgagaaga	32280
gcaactccaa	gacacataat	tgtcagatac	acccaaagat	aatgaagga	aaaaatgtt	32340
agggcagcc	gagagaaaagg	tcgggttacc	tccaaaggga	agcccatcag	actaacagcg	32400
gatctctcag	cagaaaactct	acaagccaga	agagagtggg	ggccaatatt	caacattatt	32460
aaagaaaaaa	atttcaacc	cagaatttca	tatccagcca	aactaagctt	cataagtgaa	32520
ggagaaataa	aatactttac	agacaaccaa	atgctgagag	atttttaac	caccaggcct	32580
accctaaaaa	agctcctgaa	ggaagcacta	aacatggaaa	ggaacaaccg	gtaccagccg	32640
ctgcaaaatc	atgcaaaaat	gtaaagacca	tcaagactag	gaagaaactg	catcaactaa	32700
cgagcaaaaat	aaccagttaa	catcataatt	acaggatgaa	attcacat	aacaatatta	32760
actttaaaca	caaattggact	aatgcttca	ataaaagac	atagactggc	aaattggata	32820
aagagtcaag	acccatcagt	gtactgtatt	cagggaaacc	atctcacat	cagagacaca	32880
cataggctca	aaataaaaagg	atgggagaag	atctaccaag	caaatggaaa	acaaaaaaaa	32940
gaaggggttg	caatcctagt	ctctgataaa	acagacttta	aaacaacaaa	gatcaaaaaga	33000
gacaaagaag	gccattacat	aatggtaaag	ggatcaatt	aacaagaaga	gctaactatc	33060
ctaaatacat	atgcacccaa	tacaggagca	cccagattca	taaagcaagt	cctgagtgac	33120
ctacaaagag	acttagactc	ccacacatta	ataatgggag	actttaacac	cccactgtca	33180
acatttagaca	gatcaacgag	acagaaagt	aacaaggata	cccaggaatt	gaactcagct	33240
ctgcaccaag	cataccta	agacatctac	agaactctcc	accccaaatac	acagaatata	33300
acattttttt	cagcaccaca	ccacacctgc	tccaaaattt	accacatact	tgaaagtaaa	33360
gctctctca	gcaaatgtaa	aagaacagaa	attataacaa	actatctctc	agaccacagt	33420
gcaatcaaac	tagaactcag	gattaagaat	ctcactcaa	accactcaac	tacatggaaa	33480
ctgaacaaacc	tactcctgaa	tgactactgg	gtacataacg	aatgaaggc	agaaataaaag	33540
atgttcttgc	aaaccaacga	gaacaaagac	acaacatacc	agaatctctg	ggacacatcc	33600

aaagcagtgt	gtagagggaa	atttatacgca	ctgaatgccc	acaagagaaa	gcaggaaaga	33660
tccaaaattg	acaccctaac	atcacaatta	aaagaactag	aaaagaaga	gcaaacacat	33720
tcaaaaagcta	gcagaaggca	agaaataact	gaaatcagag	cagaactgaa	gaaaatggag	33780
acacaaaaaa	cccttcaaaa	aattaatgaa	tccaggagct	ggtttttga	aaggatcaac	33840
aaaattgata	aaccgctagc	aagactaata	aagaaaaaaaaa	gagagaagaa	tcaaatagag	33900
gcaataaaaa	atgataaaagg	ggatatacc	accgatccca	cagaatatac	aactaccatc	33960
agaggatact	acaaacacct	ctatgcaa	atagactagaaa	atctagaaga	aatggataaa	34020
ttcctcaaca	catacactct	cccaaactaa	accaggaaga	agttgaatct	ctgaataggc	34080
caataacagg	atctgaaatt	gtggcaataa	tcaatagctt	accaacgaaa	aagagttcag	34140
gaccagatgg	attcacagcc	gaattccacc	agaggtacaa	ggaggaactg	gtaccattcc	34200
ttctgaaact	attccaatca	atagaaaaag	agggatcct	ccctatctca	ttttatgagg	34260
ccagcatcat	cccgtatcca	aagcctggca	gagacacaac	caaaaaagag	aatttttagac	34320
caatatcctt	gatgaacatt	gatgcaaaaa	tcctcagtaa	aatactggca	aaccgaatcc	34380
agcagcacat	caaaaagctt	atccaccatg	atcaagtgg	cttcatccct	gggatgcaag	34440
gctggttcaa	tatatgcaaa	tcaataatg	taatccagca	tataaacaga	accaaagaca	34500
aaaaccacat	gattatctca	atagatgcag	aaaaggcctt	tgacaaaatt	caacaacact	34560
tcatgctaaa	aactctcaat	aaatttaggt	ttgatggac	gtatttcaaa	ataataagag	34620
ctatctatga	caaaccacaca	gccaatatac	tactgaatgg	gcaaaaactg	gaagcattcc	34680
cttgaaaac	tggcacaagg	cagggatgcc	ctctctcacc	actccttattc	aatatagtgt	34740
tggaagttct	ggccagggca	attaggcagg	agaagggaaat	aaagggtatt	caatttaggaa	34800
aagaggaagt	caaattgtcc	ctggttcag	atgacatgtat	tgtatatact	gaaaacccca	34860
tcatctcagc	cccaaatactc	cttaagctga	taagcaactt	cagcaagtc	tcagaatac	34920
aaatcaatgt	gcaaaaatca	caagcattct	tatacaccaa	caacagacaa	acagagagcc	34980
aaatcatgag	tgaatccca	ttcacagtt	cttcaaagag	aataaaatac	cttagaatcc	35040
aacttacaag	ggatgtgaag	gaccttctca	aggagaacta	caaaccactg	ctcaaggaaa	35100
taaaagagga	tacaaacaaa	tggaagaaca	ttccatgctc	atggtagga	agaattaata	35160
tctgaaaat	gtccatactg	cccaaggtaa	tttacagatt	caatgccatc	cccatcaagc	35220
taccaagggc	tttcttcaca	gaattggaaa	aaactacttt	aaagttcata	tggaaccgaa	35280
aaagagcccg	catgcctaag	tcaatctaa	gccaaaagaa	caaagctgga	ggcatcacac	35340
tacctgactt	caaactatac	tacaaggcta	cagtaaccaa	aacagcatgg	tactggtagc	35400
aaaacagaga	tatacatgaa	tggaacagaa	cagagccctt	agaaataacg	ccgcatatct	35460
acaactatct	gatcttgcac	aaacctgagg	aaaacaagca	atggggaaag	gattccctat	35520
ttaataatgt	gtcagggaa	aactggctag	ccatatgtag	aaagctgaaa	ctggatccct	35580
tccttacacc	ttatacaaaaa	atcaattcaa	gatggattaa	agacttaat	gttagaccta	35640
aaaccataaa	aaccctagaa	gaaaacctag	gcattaccat	tcaggacata	ggcatgggca	35700
aggactttat	gtctaaaaca	ccaaaagcaa	tggcaacaaa	agccaaaatt	tacaaatggg	35760
atctaattaa	actaaagagc	ttctgcacag	caaaagaaac	taccatcaga	gtgaacagggc	35820
aacctacaaa	atgggagaaa	attttcacaa	cctactcattc	tgacaaagg	ctaataatc	35880
gaatctacaa	tgaactcaaa	caaatttaca	agaaaaacaa	acaaccctgt	caaaaagtgg	35940
gtgaaggaca	tgaacagaca	cttctcaaaa	gaagacattt	atgcagccaa	aagacacatg	36000
aaaaaaatgt	catcatcact	ggccatcaga	gaaatgcaaa	tcaaaagcac	aatgagatac	36060
catctcacac	cagctagaat	ggcaatcatt	aaaaagtca	gaaactacag	gtgctggaga	36120
ggatgtggag	aaataggaac	acttttacac	tggtgtgg	actgtaaaact	agttcaacca	36180
ttgtggaagt	cagtgtggcg	attcctcagg	gatctagaac	tagaaaaacc	atttgacc	36240
gccatcccat	tactgggtat	atacccaaag	gactataat	catgctgcta	taaagacaca	36300
tgcacacgta	tgttattgc	ggcattattc	acaatagcaa	agacttgaa	ccaacggaaa	36360
tgtccaacaa	tgatagactg	gattaagaaa	atgtggcaca	tatacaccat	gaaatactat	36420
gcagccataa	aaaatgtga	gttcatgtcc	ttttaggg	tatggatgaa	attggaaatc	36480
atcattctca	gtaaaactatc	gcaagaacaa	aaaacccaaac	actgcata	ctcaactcata	36540
gatggaaatt	gaacaatgag	aacacatgg	cacaggaagg	ggaacatcac	actctggg	36600
cggttgtgg	gtggggggag	gggggaggg	ttgcattgg	agatatacc	aatgtctagat	36660
gacgagttag	tgggtgcagg	gcaccagcaa	ggcacatgta	tacatatgta	actaacctgc	36720
acattgtaca	catgtaccct	aaaactaaa	gtataataat	agtaataata	ataaaattaa	36780
aaacagaaaac	atttatatcta	tctcttgg	ttcatagcaa	gtcatcaata	catgtttcta	36840
aattgagaaa	gaataattac	aattgaaggc	acatagtgtat	gaaagaaaaa	ttgttgattc	36900
tgacatttgg	agctgaaaat	atthaatagc	tacaatctt	aaaagtca	atttgaaatg	36960
aaaagttctt	ttatatatgg	ttgggtgatc	tgttgat	tgttttattc	ttcttaaagc	37020
aagggtgtct	tgttctttgg	acatatttc	agtgtgtga	tttgttttc	tttgtctcg	37080
tctctttctc	tctgttgc	tctctgtatc	actgccttc	tctcaattcc	tat	37140
tttggtttag	cttatattgt	gtcttacaga	cttccat	gttggcctca	gccttacata	37200
ctttcaggac	aatctcagta	cagacaaggc	tgctgcttgc	ttttggaaat	gactcctt	37260
gcctgaaatg	gctcaagtgc	acagcagg	tatgtgtgtt	ttttggat	gactcctt	37320
gtatgttcat	gtatgtgtct	ggtgaggt	caatcata	acgagaataa	ttttgtct	37380

acaatagcta ctaccattta ttttactttt tgcaagtaaaa tttttttctt acccactgct 37440
ctgttcatta acaaacttct agttaggtaa gaagggttta ttaactaaga gtttgttaat 37500
gatagttat cctgttcaga tgtgagaaag aatgttaaaa ctagtgtcta ctctattagg 37560
agatacatta gatgtatctca catgcacaat tcctatgtcc atttgcaaa catattact 37620
gcactgccta ttataatgca tagaattctg ttgggtggaaa aatacaagtg tgtccagaca 37680
tttgcgact ggtttgcata gctaataattg ttatatgtt aaagtgcata atttgagatt 37740
gactgatgca ttgtcatttt ccccatatact tgccatattt gatcaactaca ttaatttatt 37800
ttgcttgctg actgtctatt tgttctaaaca ttatttttt aaacaatcag aaacttctta 37860
gtagtatatg atactcgacataaattgttt ggactaaatt agcagttga ccatccttat 37920
gaggatcaga acaaaaatcag atcataagac atgacagaac acatatagct ttgagactaa 37980
aaccttaaaa atgtgactgt tttccctctc ttttttgagg cttcttagag ttgaaagctt 38040
catgttttt ctctttctg ttctgttat actgcctgac agaaaacatca ctttacttac 38100
catgcttatt cagtgtaccc gctttcaag gaaaatgtaa tttgtgtt ggaaaggaat 38160
ttgaaaagtc actggtcaca tcagccccaca tggataactct ggattctt ctcctatagag 38220
aaggccaaa accatttgcc ttgaccacac ccaattccca cttaaaactc tttatcttaac 38280
atctaaatgtt ctcttcttct tctctccctt ttcttatttt tttttttctt tttaccagg 38340
cttccatgtat actggggaca aaaaaagaga agataagaga tgaactagtg gctttctgga 38400
ctctcaactt gaattttttt tttttttttt gtataggggaa aatgggtt gaggaaacacc 38460
ttagcttcg atgcttaagc ttttgacttg ttgtatccctt gggtatggaa agaactctga 38520
taactttgtg gctctgtttt ctatttggtt ttgtttttaa tgattctctg gatgttaaac 38580
atcttagcaag taatttgatt tgcttaattc tgagggaaaga aaaaattcc tttgagttaa 38640
atttatgggg taaattcact tacattaaat gtaggaactt taagtgtgca gttcagtaaa 38700
gtttgacaaa tttgtacagc catgttatca caattccaaat ttcaaggct actgtgcagc 38760
tgggaagtgt gagaataggg caagtaaaga tgccacaaaag ttcaactgtcc ttaccaagat 38820
tcaggaattt ttcttgaat aaatactcac ggaatttttgc caagcttttgc attactttct 38880
agtatccaga aacagttgac ttttgacttt ttttttagtgc tctcatgggtt ttcatggagg 38940
gacagttttg ggaaggtcat tactatgtca atcagaaaatg tgattcttag atagatagat 39000
agatagatag atgatagatt catcaagatt gtagaaacag ttttttcatt tttattgtctg 39060
agagtagtaa cccattatacg ggtatatactaa taattgggtt atctatgcattt ctgctgatac 39120
atgtttgaga ctctgtatgg atatgtctt ttttgcctt gataaataact gaagagccaa 39180
atttctgtgt catagtttt tgatgaacta taaatttata taaaattttt taaaggccaa 39240
ccagttttc acagcccaattt gaattttttt cattattata tctgttatg ttttataaga 39300
gtttcagttt cttcacattt tcattgtttaa tctttgtattt gtcataatttt tttattccctc 39360
ccattgaagt gagtgtgaag tgttacaaag ataaaacaaag ccagctacca gttaaagtag 39420
taaggacaga ttttatttcg taatatactaa ttgcaatggc cagtaggtt cactgtaaac 39480
tgaactctaa ttttgttttca acagagataa ctgtgtattt taaaggtaga atgaggaaat 39540
aggaagaggc atgagccaggc ctcaagagag tcagagaagt aaaaaaatttta caaaaagtgt 39600
gaaaggagtg ttgatccatg tgaaacccac ccagtttgc caactggctc ctaccctccc 39660
ttgggagctt gtagacaaga gctcttctc cacatattgg ctgaaacacaga cagtagtctt 39720
tggcgcgc cttgagttt ctcaggcagg ttcttccagg agtactacag tcatcccagg 39780
gatatggcct tgagctgtt gccactgtgt ttgtgtttt ttcaggccaa 39840
ggttgaggcc tagtcaagaaa agtgcctcaga ggattcttc tagagtttgg tccaggagga 39900
aatctttgtt atgcgggttt tcataatggg cataatttac attttcttga caaatgcaga 39960
tggtaattt ctttcatgt gatgtttgc tggtaatgtt tttttttt aattgaactt tttatattt gaccagtggaa 40020
gctcacaact tagacttttctt tttttttttaaatttgcattt tttttttttaaatttgcattt 40080
ttgtgggaga tattttatatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40140
tagtcttctt agtctatggt ttgccttttcaatttgcattt tttttttttaaatttgcattt 40200
gttttaaaat ttgatgacac acacacactt tttttttttaaatttgcattt tttttttttaaatttgcattt 40260
attcacaattt agtgccttttcaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40320
agctatttttccat tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40380
gaattttttat ttccaccaat ttagtttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40440
aaatatttttcaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40500
acctttgttc aaaaatccat tggctacata tacgagtttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40560
tctgtacact taacctcttcaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40620
cttagtaact ctttaggttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40680
ttttcaagaa agccttcaga aaccctgtt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40740
gttaacctga ttttacttcaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40800
cttaatttatttcaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40860
atggacattt aagttagtgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40920
tgtgttttcaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 40980
tgacagcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 41040
tgctgaaattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 41100
ttgactatatttcaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt tttttttttaaatttgcattt 41160

tttttgtttc	tgagtagctt	atgattttat	gtgatatttt	taagttttat	attctactgt	41220
gtattgtctt	tatgtactg	tatgagtgtt	tttggggttt	tttattttgt	tttgttttgt	41280
ttttttagac	agagtcccgc	tctgtcaccc	agattggagt	gaagtggat	aatcttggt	41340
ctatgcaatg	tctgcctccc	aggttcaacg	gattctcctg	cctctgcctc	ccaagtagct	41400
ggggctacag	gcacgttgc	atcacaccgg	tttcactatg	ttggcttaggc	tagtctcaaa	41460
ctccctacact	caaataatct	gcccacctca	ggctcccaa	gtgctggat	tacagttgt	41520
agccatcgca	cctggcctgt	atgagtggtt	tttgtgtatt	gaatttgtaa	ttgtaatcat	41580
atcatcttcc	aaaatataca	cttaatttcc	catttgtata	ttttaaaata	acttttataa	41640
taaagcatca	tgtcatctgt	gaataacaca	gatttggttc	ttttcaattt	gtataacttt	41700
tattnagtag	gttaggactt	cagtaccata	ttgaaccgac	tacaactttt	aaataaaaaaa	41760
tagccaattg	agtaatcaca	aacacacaag	taaaatctt	agggccaata	ggatgtatcc	41820
aactgtctga	acatgggtgc	atttgtatag	ctaaatccat	taggtgttcc	actagcagac	41880
atacatcaat	tggccttcc	agcagaagca	aagtgacatc	gactaaataa	actaatgtgt	41940
tcaagtaatat	agatagataa	tataaaactt	tctgaaggct	ttctatgaaa	cataaaataa	42000
gcatgtgtaa	aactcatctt	tgagttagta	taatatacat	agctaaatct	gtattcccta	42060
accaaataatc	taagtaaattt	aatgtcacct	tctaatgtgt	ggacatctga	tcatgtgatc	42120
ccatattaaa	atttttgaga	cttgcctt	cctcatcaaa	aataaccaat	ccattactt	42180
ttgtctatcc	attcagttat	tattgagcca	gaatcattt	tagtcagaag	agtcatttt	42240
taactcttat	gttttattga	aatgttattt	catattaaca	tatgttaagaa	aacaaataaa	42300
tgaatataca	tatacataca	cacatactat	agggcatatt	gccttcaat	tgaaagctgc	42360
cattctgttt	aacaaagagc	caaattgtca	attagttta	tgcattaagg	gatgaggaaa	42420
tgagagagac	cagtattata	gtcctatatg	acccttcatt	ctatagaaat	attnaagaaa	42480
taccatttta	tttttattta	tttattttt	tattttttt	tttattttt	tttattttt	42540
tattatactt	taagttttag	ggtacatgtg	cacaatgtgc	aggttgtta	cataatgtata	42600
catgtgccat	gttgggtgtc	tgcactcatt	aactcgcat	ttagcattag	gtatatctcc	42660
taatgtctatc	cctccccccct	ccccccactc	cacaacagtc	cctggagcgt	gatgttcccc	42720
ttcctatgtc	catgtgttct	cattgttcaa	ttccaaaccta	tgagtggagaa	catgtgggt	42780
ttgggtttct	gtccttgcga	tagttgctg	agaatttgct	gaaaattgag	tttccagtt	42840
catccatgtc	cctacaaaagg	acaagaactc	ttcattttt	atggctgcat	agtattccat	42900
ggtgtatatg	tgccacattt	tcttaatcca	gtctatcggt	gttggacatt	tagttgggt	42960
tcaagtctt	gctattgtga	atagtcccc	aataaacaca	cgtgtcatg	tgtctttata	43020
gcagcatgt	ttataatcc	ttgggtatat	actcagtaat	gggatggctg	ggtcaaaaaga	43080
aataccattt	tactttaaa	ttagtgcaag	gaatcaaaac	catacagata	aaaactgaga	43140
taaggcttta	aactcaattt	gaggaaaaaa	atagctatgc	tcaaatactt	caaatttcat	43200
gggatgtaac	aaactctacat	tttgatttt	atatactat	atattgtgtt	gactgatttc	43260
tcaaagaata	accaccattt	gtgatctaag	agaatttaggg	atgccttcc	tgtcttggc	43320
agagcccccatt	ttgattcatt	tgtgcttaacc	tggactatc	agtgcactgg	aaaatacaga	43380
ctaataatgaca	catgtctagc	ttcccttattc	atctaaca	tgcagtgaa	acaagtttt	43440
aatattcttc	tttcttatttt	cagccctatt	cctaagtcaa	caattttat	actcatcaat	43500
tttataaaacg	gtccttggct	cagacctgtc	tttgattctt	agttgtttaa	tgactttcaa	43560
tagggcagac	tagtccagc	ataaaatccag	aacgtccgtc	ctccaaagga	cctggggccct	43620
gaccaataact	tctgttcc	tctactgatt	atgttgcac	tctgctgctc	tccacatct	43680
tacaaatcct	tcctttcctc	aagtcttca	gctgccacac	ctttctaccc	ctcaccctca	43740
gtaaaaggacc	tcaccccaagt	gtctatagac	aaaactggag	agttctgatg	aaaaccttct	43800
ttggtttatt	atattaaaat	tacaaaactt	atctatctat	attttcttc	ttccatattt	43860
ttgaaaggta	tggctagtt	ctctgttgt	gctaaagctc	tgatgaacac	atgtaaacaa	43920
actgtacata	aatgttgtat	agggatttca	tcttcaaagg	agaatctagc	tgttcaaaact	43980
atcaaataaa	tgttaagccc	ttgaggaaca	aatgatatta	taatttggtt	tttagaattt	44040
ctttaaaat	actctaggta	atagctaa	gaaatata	tcttgcacc	ttgctttaca	44100
aactatggtg	agttaacatt	aacaggtagg	tcatttgcta	catttgttca	ctaaaaatac	44160
aacttttct	gtgtgtggag	attatgtcc	aatcaataga	ggaaacccat	attttcatca	44220
atttcactgg	actttctat	tcatatgtt	acatttcatt	gactctaaaa	tcaatttat	44280
gcaagctct	ctgtatgtata	tgtgagagct	aaataagcaa	tttctttagg	gtgcaaattt	44340
ttgagcaata	atttttacag	tcatagctt	tttcatgaa	gaactcagag	agagtggctt	44400
gagttggcag	caaagactgc	ttttaacaat	gttctgtgt	tctctaa	tattttaata	44460
agttagtttag	tggcatataa	atggagacaa	aaaaacttca	ctctcttcc	tcaggtgcata	44520
taattgttt	tttactgttc	aaatgttgc	ttctcagaaa	aggctacttt	gaatagtcat	44580
ctaaaagaat	ctcacattca	ttgttcttca	cagaccac	ttttacttga	attataata	44640
tgtatcttc	tctgtgggtt	taaccattat	ttctcatatt	ataatgtgag	ctccatgaaa	44700
ctggaaaatt	tgttcactcc	aataacccca	ggaatttgtaa	tacacacaa	aaaatcaagc	44760
tatgcagtc	gtgtacatga	ataaacaggc	aggtaaaatt	tcaaagtgaa	aaaacagttc	44820
ttggttttgt	atattcttag	ccatataatca	agcactaagt	ttaactaatt	taatgttcatc	44880
tgacaatcat	aggttctgt	ttgtttgtt	gtttgtt	ttgagacgga	gtctcgct	44940

gtcgcccaga	ctggagtgcc	ctgggtgtgat	ctcggctcac	tgcaagctcc	gcctcccccgg	45000
ttcacgcccatt	tctcctgcct	cagcctcccg	gtagcttgg	ctacagggcac	ccgcccactac	45060
gcctggctac	ttttttgtat	ttttaggtgg	gatgggggtt	cacagtgtt	gccaggatgg	45120
tcttgatctc	ctgacacctgt	gatctgccc	cctcggcctc	ccaaagtgt	aggattacag	45180
gcgttaagcca	ccgcgcccgg	ccgacagtca	taggtttaa	gtgaaaatgc	taccttaatt	45240
tttttcttct	gttttgtatc	tgaaataata	agatatcaga	catggaggga	aactgaagat	45300
tggacttcct	acatttattt	atttaaatga	tgctaatatt	aatatctgtat	attcatttct	45360
ttgttttttt	tgttccctca	caaaccctga	aaataattac	caaatgtaga	attcaccaag	45420
aagacaaga	taaatgcatt	aatgtataaa	tttataatgt	tatatatata	agagacagaa	45480
atagagaatg	tgcctgtgt	tgtgtgggtt	ttagatacaa	aaagagacag	aaaacacaca	45540
catataaaaga	gaaaacgttg	tgcagaaatt	ctgatgtgtt	tattaaaaaa	agtaattgtt	45600
gcctagatga	atagatgaat	gagtgaatga	taaatggatg	aaacaaatgc	caaatctgga	45660
ttagagagaa	tcctcacatt	ctttgtcact	ttcagtttc	aagaataaag	aagatgttgt	45720
ccccaaacca	caccatagtg	acagaattca	ttctcttagg	actgacagac	gacccagtc	45780
tagagaagat	cctgtttggg	gtgtccctgg	cgatctacct	aatcacactg	gcaggcaacc	45840
tgtgcatgt	cctgtgtatc	aggaccaatt	cccaactgca	aacacccatg	tatttcttcc	45900
ttggtcacct	ctccctttta	gacatttgct	attcttccaa	tgttactcca	aatatgctgc	45960
acaatttctt	ctcagaacag	aagaccatct	cctacgctgg	atgcttcaca	cagtgtcttc	46020
tcttcatcgc	cctagtgtatc	actgagttt	acttccttgc	ttcaatggca	ttggatcgct	46080
atgttagccat	ttgcagccct	ttacattaca	gttccaggat	gtccaagaac	atttgcatct	46140
ctctggtcac	tgtgccttac	atgtatggct	tccttaatgg	gctctctcag	acactgctga	46200
ccttcactt	atccttctgt	ggctcccttg	aaatcaatca	tttctactgc	gctgatectc	46260
ctcttacat	gctggcctgc	tctgacacacc	gtgtcaaaaa	gatggcaatg	ttttagtgg	46320
caggctttac	tctctcaagc	tctcttca	tcattttct	gtcctatctt	ttcatttttg	46380
cagcgatctt	caggatccgt	tctgctgaag	gcaggcacaa	agcctttct	acgtgtgctt	46440
cccacctgac	aatagtcaact	ttgttttatg	gaaccctctt	ctgcatgtac	gtaaggccctc	46500
catcagagaa	gtctgttagag	gagtccaaaa	taattgcagt	cttttataact	tttttgagcc	46560
caatgctgaa	cccattgtatc	tatgcctac	ggaacagaga	tgtaatcctt	gccatataac	46620
aaatgattag	gggaaaatcc	ttttgtaaaa	ttgcagttt	ggcctgtgtt	tatttgtaat	46680
ccctaagtg	ctgtggggta	acaaactgaa	atggaaaaac	ctagtgtagt	tattattnaa	46740
cagtagggc	tcttagtaac	cacttttagt	tcttctcaaa	attaacactt	tgaagattta	46800
gtttttaaaa	ataaaaaagct	taatgttga	attaataatg	tttttattgt	cagagattct	46860
atgaatataa	attgtttgtt	ttctaataat	tctgtatgaa	aatactggc	ttacttactt	46920
gggggttaaac	ttttaactaa	tttagctga	gagaaactct	gaaaatccaa	cccctttgaa	46980
tattactatc	catagagcca	tagtgaggct	ccagataggg	tggctattt	tacatttttta	47040
aaaaaaatgt	gccggggcatg	gtggctcaca	tctgtatcc	cagcactttg	ggagccagag	47100
ataggtggat	cacctgaggt	caggagttt	agaccagctt	ggccaacatg	gtgaaaccccc	47160
gtctctccct	aaattacaaa	attagcttgt	tgtgggtggca	catgcctgt	atcccagcta	47220
ctcaggaggg	tgaggctgg	gaatagctt	aacctggag	gcggagggtt	cagtgagtc	47280
agatcacgccc	attgtactt	ctccacttgg	gcgacaagag	cgaaactcca	tctccaaaaaa	47340
aataaataaa	taaataaata	aaaagaaaaac	gaaaaaagaa	aaaaattgtt	tctacttcaa	47400
tgtgtgtct	tgaacaatag	ttctgaaaacc	tgtgaactt	ttgtatcac	tgtgcaaaaaa	47460
atgtaaatat	atgaatcacc	tataagttaa	attttctgtat	agtattctca	aaacccacac	47520
tgtaaacatc	gtgagaagct	tcttatttgtt	tcatgtggta	ataccagtta	ttgtccttcaa	47580
ttctgaaaat	caactctcaa	cacccttttt	tttttttga	gttttcaagt	gcatggatcc	47640
atttctatca	tcaaaatact	tcaattgtat	tccacgtat	ttttccagtg	aagtacctt	47700
gacttaggaa	taatttagta	ggaatcaata	gggtttatca	attgtatattt	gcatatgtat	47760
aaatgactt	gttagtggaa	aaaggggtct	tgatccagat	cccagcaggg	gttttcttgg	47820
tctacagagg	aaaaatttca	agataagctg	caaagtgcag	tgagaagaga	tagttcatta	47880
aatgtactc	caaagagatc	atttattaaa	ggctattgca	ttacagagta	aggtgttccc	47940
agaaagcaag	tggaggaatg	cacccttttt	aagttttttt	aatggctttt	tatctacgt	48000
cagactaaag	taagttgtgg	ctacatgtgg	gtaggctgac	ggcgtatctt	attttatcat	48060
tgtgttgatt	taaagaaaaac	gatccttgcatt	atttttagtgt	gtgcataact	attattatca	48120
tgataaaaagc	atataactt	atgagaattt	ggacacctag	gttcttgc	tgcatttata	48180
gcgtgttctt	gtaggtttt	ttaggctgtt	tcctaaacta	taagcatctt	atgaacatgg	48240
gtcatgacgg	gccaggactg	tgccttgcatt	acctaagac	agagtgtatt	ttaaaaatgtt	48300
gtcacccctgg	atctcctaga	ctcctgcattc	tttaacaccc	tgaaacagta	taaaaattaa	48360
tagggcgt	cgtatctttt	tggagggaaat	agtgtttttt	ttgcttcaca	caactgtctg	48420
tggcacat	tttttttaca	ttagggttatt	ctagttctca	aagctactac	agggaaaaatc	48480
aatcaca	aggctaaata	aagatacaca	ttaatcaatg	tattattnat	tcaaccagat	48540
gactgcata	taatacatac	atatttagcc	ctattaaatt	gcctttatg	tttgacat	48600
taatattttt	caactctgtt	ttaactaaca	ctatacagaa	aatgtacag	aaaatgtctt	48660
tatgatatac	actactgtat	ctactataaa	atgttcaagt	ggtttccag	aatattaaac	48720

agatgactaa cagttttgt	aaaattgggt tatattctta	agtgtcagtt gggtgaataa	48780
gtctggttct atccaaatct	gcttttagaa acagtaactg	aaaaagaaaag aaggcaattt	48840
tgggcagcaa caatttaata	tttacctttt agcactattt	gtgtatttgc ctgggtttgg	48900
caatggcagt ggcacaacag	aagcaggta tctgttagaga	caagagtgaa gcagaagagc	48960
tgggactga gtgacacagc	gtccttactt ccttttagag	gcccagcatc atttcatctg	49020
tgggatattt ctatctta	tgacaacatc atttaatctg	tgggatagct ctatcttata	49080
ttcttttatac ttccttttt	ggttaggcaa agcattgtgt	tcaatttcta gcaagtagag	49140
tggctctgata ctctatttga	aaactgtattt ttagaattgg	tacaaacaca ctctatattt	49200
tgcctaaata actgtatact	tccttttattt ttcttagttat	gagaattgct gtggcttgca	49260
gcaacccaaac ttaaactttg	acaccttgt aacccaaatg	cttgacactgt ttgtatctc	49320
tttttttaa aatatgtatt	tatttattctt atttgttaatt	taattttta atttgcaag	49380
cccttggta aaagttatct	cagaaattctt ttggagatag	aatagaacat aaatggacat	49440
gaatagaat agtacaatt	gatgtaatca tagtgtattt	ctaatttttag gattactgtt	49500
ttttcttac ttaaatgaaa	atgaagtagt gagagaatta	ttatcttcat ttacaatatt	49560
tggagtctta gacacagaga	gatgaaatga ctgcctaacc	gtcatttggc cactcagtag	49620
cacaaaagca gatttcttgg	ttttaaattt tctcttttcc	taattcatac tgaagtgggt	49680
attaagccat tcttcaacac	acttcaatg ctggggggaa	tattggagac tatgaattaa	49740
taagagtgtt actatagatt	aaatagtggt tattatgtga	tagtgagatg tattcctaaca	49800
acagctaattt tgaattttcc	ctgacagtca ggctaatttg	atcattata tgcatactaa	49860
ttaaaattgtt cttaatatcc	atgtgaaggg agaaaatttt	gatatatggg aaacctttt	49920
gcttcttagg ctgcattttt	ttaatatattt ttcccaatta	taaaagcaca catatttata	49980
ttaattaattt tccaggtaca	gacaaatataaaaacataaa	atttataag gacctgctat	50040
gacaaggaat aagttttgtt	tatattttgtt tatatttctt	ctcaataaaat gtatatgact	50100
acatgttact ataatgcaca	aaacaaatataaagttactc	accatataat aatatgaata	50160
cagattttca tttgatacta	taatgtatgc acctttacat	agattaaata ttatgttac	50220
atttttaatgt tcatattttt	attagcttgc tgatcctaa	tacaatatttata atacatgtca	50280
aatgtagaaaa accaatgaac	catattggca tacataagaa	atagaagcaa tgcttacttg	50340
taatcaacta cctgaaaaga	gttgctctca gcattcacat	gtgtttattt ccatttgatt	50400
atcttgcaaa tcatgattcc	atcatatcaa tttatcttag	tattttgtatg accgacgtat	50460
ctgtacgtct ctctctaattt	gttcccttttgc ttatgttccc	agaagtgtaa ttgctgaatt	50520
aaatgttttg aacatttttta	aaggcttgc taaacatttc	tatggataaa ttatgttacca	50580
gaaagcttgc aacaaacctg	taccactaat gttgaaaac	aatcatttttataactct	50640
atcagcagta ggagcttttgc	tttatgttta tatgctactt	tgtatctca aaggaagttt	50700
tagaaacaaa ttctgttttta	atttgcacag attagaaattt	gaagaatata ttacaacaaat	50760
aaaggctaca tataataattt	ttctttcatg ttatgttgc	gtcttacatg aagaatattt	50820
gtatattttat ggagaaaaat	tatattttgtt ttatgttgc	gtttaactt tttcttattgc	50880
cacaatttggaa taaacaaagca	cttataatggg gttattggct	ttcattgcaa tattttatgg	50940
gaatattcatg aattatcctg	gatataattt ttttgcattgt	tttaatgttgc taaattactt	51000
tttgtccctc acatttttaa	ccaaacttcc ccaagtgta	ttaaggactg ctttccact	51060
gattttgtat gtcaccttta	tttataccctt agtcaatcta	tttctgccc tttccactgt	51120
ttttttgttc tgaatttataa	atataatttttta aactaaattt	ttatcattaa attgaacata	51180
tatattttat tctggtaag	atattttgtt ctcaattttt	ctctttaggat attgaaaatc	51240
agatactttt agctatccaa	atttttaactt cagaaattat	aatattatgtt tcaattttttag	51300
aaactgtatg tataatgttgc	taattttgttgc tttatgttgc	catacaatac ttatacattt	51360
taaagcaattt aaaaaataag	gaaaattattt ttactgtcta	tggtgccat agggagaact	51420
gtttaaggat gaaatattttt	gaatgatgtt tttttttcc	agaaaatgtcag gaagatgtc	51480
gtgtctggta agtggggaga	aaacacccaa ctttggtata	ttaattctat ggattttata	51540
acactgttatac acaaataactt	atgttttccac cataccaggta	tatggtaaaa tcatcatgt	51600
acattttttt tcaacaaata	tctctttaata cagtctttt	ttcttacaaa gaggtttttt	51660
tcccagcaat tcccccaaaa	agaactgaga gggccacata	tgttttccaa ctatgagtc	51720
cttcaacac tgaacaaaaaa	aggattaaatc agaaaagaaaa	taaatatttta aattttaaata	51780
atcagagaat aatataatctg	tgcttcataa cagtggggaaa	aattttgtat aaaaagagta	51840
agcaaaattgc tttatgttgc	agataaaaactt attttaaat	aagaggcacc ctacaatttg	51900
gcaatgtatt tgccacaaac	aagataaaattt ggtaaaattt	ttgaaataca agaaatttaat	51960
aaaaataattt tgaagcacta	aagaatcaat atataaaat	caaagactat aagttaaaaaa	52020
ttgattttttt taaaatgttgc	ctaaaattaa cacacaaaat	attagcaacc aaaaatatgt	52080
tgttgcattat ttttttttttgc	aaaagagctg atataatgtt	gcactttcat	52140
acattgtattt attaagggtt	taaattttttt ctgaaaatata	taataggttgc	52200
gtacaaaatgtt aatcttgc	ttttgcatttaaaagtaaca	gcacttattt tttttcttac	52260
caacctaata ttattatcat	attaaggaga caaaatttttgc	ttatgtttttt attcagctat	52320
ttcaattttt agaatactact	gaaggcataa gaaacttgc	tgaagatgtt catgggttgc	52380
ttttttttttaaataatcaaga	caaagtaaac acatgtatgt	ttaaataacaa atagttccctg	52440
tttggcatgg taccatgtaa	cagaaacccctt tccatactgg	aaacgtgtct tagtttttgc	52500

ggttcttagt	tcctacaaaa	tcacaaagtg	agaactgcct	gcatttaata	actggtaaaa	52560
tacaggcatc	caccagatgg	gatagaagct	gtctctaact	aagtagaaatg	tcattagcct	52620
tgatcatgaa	ggaaagtgt	aagtacaggt	gaaacgtttt	ttgtaataga	aagttataca	52680
gccattaaag	tgaggatgtt	cctagttaaag	aaaaaggtaa	ttcattttat	gtaatataatt	52740
atgtgcaaag	atatatccta	tgctgtatac	agtgtgatat	cttgtaaaat	attagggata	52800
aggggaaata	aaacttgtta	ttaatagcta	tgtccatatt	taaatagtca	ttgtctattgt	52860
gagaaaataa	ggcaacaaaa	actgtgtatt	ttcagtttag	tccagatctc	aggatgtata	52920
attacaggat	acttattttat	taatatttga	atthaataat	gaaagtaagg	aaaagcaagg	52980
aaaaagcaag	gtaagaaaaa	gaagagggtcc	ccaattctgg	tttccagcac	agaatgacca	53040
atgaacaaact	tctcagagaa	aaatgccat	gtaaaaattt	cgtaacccat	gggtgaggtg	53100
gggcacactc	ttggagcata	aaaactggaa	aaaaccacat	tagaattgtt	aagaggagca	53160
gtttcacttg	cttcttcc	agccaaacac	agtgtcacac	tgaaagagat	tccctgggccc	53220
tgtggttcc	ccagtgaaga	aaagagagcc	cacggcagac	atccagcttc	cctatgttcc	53280
agagaacttc	ccaagaggcc	cattctgtct	aaccttgtgg	gtaataacgg	aagacttggc	53340
aaagcttggc	caccccgatc	agctcataac	aaagtttaagg	ggttagtctt	acggcaacca	53400
gcacccgatc	ttgggggtgg	catcatgttc	tagctggcag	ctttatgtga	ccagagagcc	53460
cagccactag	cattgctcac	atgtggagct	gagttggcag	gcccttcgg	tcaggaggca	53520
tgttgaacag	ttctgtctgg	ctgtggtacc	agtcagtaag	ctggaccagc	tgcagagcac	53580
agcctagagc	accacccaaac	cacaaggcac	aacttgcagc	ccacccact	caggttagaga	53640
tccaagccag	caacccact	caaccgctgg	gcatagcctg	cagtcccatc	tgacgacgga	53700
gtctggtcaa	ttatctctcc	aaactacaga	gcacatccag	taaaacttcc	caattatgg	53760
tcagcttgc	gccccaaaaa	actgcagggaa	atagccagtg	gccctatctg	gccaagagat	53820
ctggtcagtg	atctcaccta	aatttggagc	aaaggcagtg	accccatcaa	tctgtggacc	53880
acagccagca	gccccactcg	aatacatacc	acagtgaatg	gtcttacctg	attatggagc	53940
ccagtcagca	gccctgagtg	atttggatc	ctagccagca	tactcaccca	cctcagagca	54000
caggcagcag	tcagaggta	tgtgactaga	gtcagcttcc	aaattcatca	agtcctgggt	54060
ccaccccttta	ctcccttatga	aatctcaggc	aaattattta	acctctactc	ctatgccccca	54120
atttggttgt	ttgtttgttt	tcacagttact	tatcatcacc	tgccatatat	ttgtttgtgg	54180
gtgtattttgt	catctatctc	ctccagcaaa	aaacatttac	aataaagttag	tttttaattt	54240
tttttgagca	acagctatata	tgagttatata	agtggtctata	ctaatttaca	tttttgccaa	54300
taatgtacga	gcattcctt	ttctgtgtat	ccttgc当地	ctctgttcat	ttttttgtct	54360
ttttaatgac	aatcattaca	aatttgggtga	ggttataactt	cattatggtc	ttgatttgca	54420
tttcatttat	gattagtgaa	ttaaacccctt	tttcatata	cctgctgaca	atttgaatat	54480
cttctttttt	aagtttttga	taaaatccaa	catcttca	taataataat	ataaaaaaaa	54540
tactcaacaa	actaggcatt	gaagagacat	accccaaaat	aataagagcc	atctatggca	54600
aacccacagc	caacatcata	ctaaatggga	aaaagctgaa	agcatccac	ctaataaaatg	54660
gaacaagaca	aagatgtcct	ctctcatcat	tcctatttca	catagttactg	caagttctag	54720
ccagagcaat	aaagccagag	aaagaaatata	aaggcatcca	aatagaaaaa	gaagtcaagc	54780
tatctcttctt	cactgataat	atgattttat	acctagaaaa	ccctaaagac	tacaccaaaaa	54840
agctcctaga	tctgataaaac	aacttttata	aagtttcagg	ataaaaaatc	aatgtataaaa	54900
aatatttagca	tcttataata	cccataacat	tcaagctgag	agctaaatca	agaatgtata	54960
cccatatttaca	atagccacac	gcacaaaaat	aaaacaccta	ggaaaaacaac	aaaccaagga	55020
ggtgaaagat	ctctacaagg	agaactacaa	aacactactg	gaagaaatca	tagatgatata	55080
aacaaatgg	aaatcatccc	atgtttatga	attagaagaa	tcagaatgt	aaaacagctaa	55140
tattgcctaa	agcaatctac	agattcaaca	ctattgccc	caagctaata	acactattta	55200
catagaattt	gagaaaaactt	gcctaaaatt	tatatggac	caaaaaagag	actgaataatg	55260
caaagcaatc	ctaagcaaaa	agaacaataac	tggcagcata	aaataaccca	acttcgaact	55320
atactacaag	gctacagtaa	ccaaaaaaagc	atggtactgg	tacaaagaaa	gacccataga	55380
aaaaatgaaa	cagtgaaccc	agaaaataaa	tcagacatct	acaactgtct	gatcttaat	55440
agttgtatag	taataaaca	tagggaaaga	actacctatt	caataaaatgg	agccaggata	55500
actattttagc	catgtgcaga	agagtggaaac	tggacccata	cctgtccatca	catataaaaa	55560
ttaactcaag	atggattttaa	gactttaata	taaggcctaa	agctataaaa	atcctggaa	55620
aaaactgagg	aaataccact	ccggacattt	gcctcagcaa	aaaattttatg	aataagtctc	55680
cagaagcaat	tgcaataaaaa	ataaaaaattt	aaaaatggga	tgttaattaaa	ctaaagagct	55740
actgcataga	aaataaaatta	acagagtaag	caatctaaag	aatgaaagaa	aatgttcaaa	55800
aatgataat	ccaaacaaaaa	tctaatatcc	agaatataata	agaaacataa	acaattcaac	55860
aggcaagaaa	cctccccctt	aaaccattaa	aacacatgg	cacaggagg	ggaacatcac	55920
acactggggc	cttttgtgg	gtgggaggct	aggtaaggga	taccattagg	agaaatactt	55980
aatgttagata	acggattgtat	gggtgcagca	agccaccatg	gcatgtgtat	acctgtgtaa	56040
caaacctgca	cgttctgcac	atgtacccca	gatcttaaaag	tataaaaaaa	aataggcaaa	56100
agacataa	agacacttct	caaaatgtaa	catacatgtg	gccaaaaat	atattaaaa	56160
tacagtatca	ttaatcatca	gagaaatgt	tatcacaact	acaatgagat	aacatctcac	56220
accggcgtcaga	atggcttatt	ttaaaaatgt	aaataaataa	cagacattgg	tgagggttg	56280

agattatttt	tcatccttt	actttgagcc	tatgtgagat	agatctcttgc	aaaattgcag	60120
atggatggat	cttatttttt	atctaacttg	ctactctagt	atatgcctgt	ctttataggt	60180
aaatttcttg	aagacagcac	atacttgagt	attgattttt	gtttgttgc	ttttcttttg	60240
ttttaacttt	ttttccattt	agcagtact	cagcgtctt	taattttaga	attttgcata	60300
ttaacatgg	agagtattat	tgttaggtaa	gagtttacta	ctgccatttt	gttatttctt	60360
tgcttgggt	tttctaaatc	ctttatttct	tgtgttctat	cttactgtat	tactctgtag	60420
ttaagtaatt	ttcttttagaa	gtatgttttgc	attccatgct	atgtttttgt	gtatctgtca	60480
taggtttttg	ttttgtgggt	accgtgaggt	ttacaaaaaa	aatctatagt	ataacagttc	60540
gttttgaaca	gataatgact	taattttgat	tgcaaaataaa	agcataaaaa	accaactcta	60600
cacttttaat	atgccttaca	cacatttga	cttttgggt	ttcaatgtac	atgtttttat	60660
attgcctact	tctaaactgt	agttatttgc	ggttttataa	gttttgcattt	ttatttcttca	60720
tacttatgtat	acaagtgggt	tatttaccac	aattatagta	tgacagtatt	ctgattttga	60780
atgttctttt	acgagggaaat	tttatacatt	ccaatgtttt	cctgttacat	gttagtaccc	60840
atttctttca	gactgaagaa	ctctctttag	catttcttgc	aaaacaggc	ttgtgttgc	60900
gaatttcctc	agctttgtt	tttctgcgaa	aaaaaaatct	ttatttcacc	ttcatgttgc	60960
aagaataact	ttaatgcaca	taatattttt	ccttggaaaga	tttttctct	tatcatcttgc	61020
agatagcat	accatttcct	tttgacctgc	tagttctgt	gagaatctg	ctgaaagcca	61080
tattcaagct	ctgttcttac	ctgttcttac	attgctaaac	aaaattacct	tagcctgagt	61140
agtttataaa	gtgaagaggt	ttaatttgcaca	cacagttcca	caggctgtac	aggaggcatg	61200
gctgggaggg	ttcaggaaac	ttacaatcat	gatggaaagg	gaaggggaaag	caagcacatc	61260
ttcacatggc	gacagaggag	agagggtaaa	ggggaaagg	ctacacactt	ttaaacaac	61320
agatctcatg	agaactctat	cataagatag	cactagggga	atggtattaa	accatttagaa	61380
aacaccccg	tgatccattt	acctcccacc	aggccccacc	tccaaaactg	ggggattaca	61440
attcaatata	aggcttaggt	gggaacacag	agtcaaacc	tatcattctg	tcccggtccc	61500
tcccaaattt	catgttctt	tcacatttca	aaacgcattt	atgtttccc	aatagttccc	61560
caaagttttt	actcatcattt	gcacaaactt	aatgttccaa	ttccttcatgt	ccaaatgtcc	61620
aaatgtccaa	aatctcatct	gagacaaggt	ccacaggtaa	gcctgtaaaa	ctaaaaacaa	61680
gttatttttt	tctaagat	aaggggat	caggctgtgg	ataaaatgtt	ctgttccaa	61740
aggaagaaat	tggccaaat	aaaggagcta	catatgcct	gcaagtccaa	aatccagaag	61800
ggcagtcatt	aaatctttaa	gttccaaat	aatcttctt	gactccatgt	ttcacatcca	61860
ggccacacta	atgcagggtt	tgctcttca	aggcctgggg	caactccacc	ctcatggctc	61920
tgcagggtgc	agccctgtt	gctcttca	caggctgtt	ttgagtgctt	gtggctttc	61980
cagatgcatt	gcacaagctg	tgagtggagc	taccattctg	gtttctgaaa	gatagtgtcc	62040
cttttcttac	agatccattt	cgcaggacc	agttggaaact	cggtgtgggg	gttccaaaccc	62100
cacattttcc	ctctgcattt	cactagtata	ggctctctgt	gaggattctg	ccccctgcagc	62160
agacttctgc	ctggatattt	aggcatttgc	atacaaccc	taaaacttag	gcaaaagttc	62220
ccaaggcttta	attcatgtt	tctgcacacc	cacaggcc	acactatgtg	gaacttgc	62280
aagcttgggg	cttgccacc	ctgaagcaat	ggcccaagct	gtaccttggc	actttttat	62340
cctgacttgg	gttggagtgg	ctaggacgca	ggtcatcat	ttctgaggct	acacagagca	62400
ctgcagacat	gggtctggcc	cccgacatca	tttttccata	tttgcctct	gagcctgtga	62460
tgggagtggc	tgttataaa	gtctctgaaa	tttcttggag	gcattttccc	cactgtcttgc	62520
gttatttttt	tgtagctt	tttattttat	gcaaaacttct	gcagcatttgc	atttccctt	62580
ataaaaatgg	tttttctttt	ctaccat	ttcaggctgc	aaattttca	aactttcatg	62640
ctctgctt	ctttttaata	taaatttca	ttttagat	tctcttgc	aatagagaag	62700
aagccaggat	aatttcttgc	tgctttctg	tttagaaatt	tcttccacca	gatacctgaa	62760
ttcatcttc	tcaatttca	agttccac	cttctctag	cagggcaca	atgcaccat	62820
tcttcttgc	aaagcatagc	aaggttacc	tttactccac	ttccttataaa	gttcttcatc	62880
tccatcttgc	accacccat	cctggactt	actctccata	ttacttattt	tattttgc	62940
accaccatc	aacaagtctc	taggaat	caaacttct	cctcttcttgc	tcttcttgc	63000
agccctccaa	actgttccaa	cctctgtt	ttaccaatt	acaaatgtca	ttccacat	63060
ttgggtatct	ataggaatgc	cccacttctc	tggttaccat	ttttagtatt	agtccatttct	63120
cacatca	taaagaacta	actgagattt	agtaatttt	gaagaaaaga	gtttaatcg	63180
attcaccgtt	ccttaggt	tataggtac	atgggttgc	aggcctcaga	aaactaaca	63240
tcatgggt	aggccaaagg	aaaacaagc	acatggcag	aggaaaaaa	cagagcga	63300
ggggaaagtgc	tacacactt	taaataacag	atattgtgag	aactctaaca	caaggcaca	63360
ctaggggaat	ggtgcataac	cattagaaat	caccccccatt	atccaaatcac	ttcccccac	63420
gtcccaccc	cagcagttag	agacggc	tttagat	atttgggtgg	agacacagag	63480
ccaaaccata	tcagtc	ttaaatgt	tattttctc	ttgctgattt	gaatattt	63540
tcttgc	ttttttcc	taatttgc	acagtgttac	ctgagatatt	actgtttgt	63600
ttgaatttgc	ttgggtgagca	atggtctttt	ctacctagat	gttattgtgt	ttctccagaa	63660
ctgagaaatt	ttcagttt	atttcaataa	gtatgtttc	taggcctttt	tatctctcat	63720
tccatttgc	tatttcaatt	atgaaaatgt	taatgtactt	gattgtgtcc	cataattttc	63780
ataggcctgc	tttactttt	ttttttcc	ctcctctgag	aagttat	tttcaatgtt	63840

atcttcaagc ttactgattc	tttcctctgt	catcaaaaact	atgggcacgg	tggctcacgc	63900
ctataatccc agcactttgg	gaggccgaga	agggtggatc	atctgaggtc	aggagtttg	63960
gaccagtctg gccaacatgg	tgaacccca	tctctactaa	aaaaatacaa	agaattagct	64020
gtttagtcatg atggcacct	gtaatcccag	ctacttgcg	ggctgaggca	ggagaatccc	64080
ttgaacccgg gggcagagt	ttgcagttag	ccaagatcgt	gccatcgca	tctagcctgg	64140
gcaacaagag cgaaactctg	tctcaaaaaa	agaaaaaggaa	tagaggaggt	tggaaagagg	64200
gaggcaggga gtcagcgtca	tagcaatggg	tgagaaagat	gaaagatgtg	tggccattgc	64260
tggcttggaa ggtagaatgc	gggcagccctc	caaaaacctgg	aaaaagcaag	aaagtagatt	64320
ttccctgaga acctccaaaa	agaaagcagt	cctgcccaca	ccttgcattct	agccccagtaa	64380
gaagcatttt gggcttgg	catccagaaa	tgcaagacaa	aatttgtgtt	gttttaagggg	64440
attaaatctg tgatagttt	ttatggcagc	aacagaaaaaa	gaatatacct	gtcaaaaactg	64500
gtttcccatc ttcaaggct	tcccctctgc	ttgcccgtcca	gtccacagct	ccacttgcag	64560
gtgcagtaca gttggcttc	gagcagggtct	tgaggatggaa	attcagcatg	taaggaagaa	64620
gtctaaagtc cagacgaacc	ttgaacacccc	cttaattgccc	cctccgtcg	agagcaggga	64680
ccaggatcca ctgagtgtag	agcagagcaa	cattcctcaa	ggtgtcctca	ccccccaggca	64740
ccctcccaaca gggctcaacc	agctaaccctc	tttccaccct	ttgaattgtc	tctctctccc	64800
ttcaggcata taattaaatt	tgcataaaatt	caaagagcat	atgctcaat	cccattgtaa	64860
tttcagattt gtttttattt	cctgagtttgc	ctaaggaaat	aataaaaaaca	aaacccctt	64920
aaaaaaaaaa aagatttcta	atgagttttt	aagttcagtt	tttgtattct	ttatttctgg	64980
gattaatatt tttaaacatt	gtttttattt	ctctttccaa	tttcttattt	tcttcctgaa	65040
ttatttctt tttttttag	ttttctgttt	tcttataatt	tctttttta	acattttct	65100
ttcttttaat atttttttaa	attccttattt	tcctcagaga	tctttgttct	gacctgtaat	65160
ttcttgaact tcttaagag	gtttatgtcg	aattcattgt	cagacattta	atagatctt	65220
aattgttcag ggtccattgt	tgagcctctg	ttgatttctt	ttgggtgtgt	catatttccc	65280
ttggtttta taatcttcgc	atatttatgt	cgatgcctgt	gcatttgaag	agagagacac	65340
ttccagttt cgcaagggttt	ctttgttgg	cttagatctt	tagttcttag	tattttttt	65400
ttttttttta agacagagtc	ttgctctgtt	accatgaat	gcagttgtgt	gatcttgggt	65460
tactgcagcc tctggctccc	aggttcttcc	agcgattctc	ctgcctcagc	ctcccaagta	65520
gctgggatta ccggcataca	gcactacgccc	tggtcaattt	ttgtattttt	agtagagatg	65580
aggttttacc atggggcca	ggctggctc	gaactcctgg	cctcaggta	tccacccggcc	65640
ttggccaccc aaagtgcgt	gattacaggt	gtgagccact	gtgcctggcc	agtattca	65700
cttaaatgtt gacttgtgt	tgcttctgt	ggggtaggtct	tatagtgacc	actgcaacta	65760
aatttctttt ctgtcattgt	ttccttagtgc	ggggaaatct	tatttgtgc	actgaagctt	65820
aaatactgtt ttgggactat	attgctgtgc	tgtcattgtt	tccaaattccg	gaaagttatt	65880
aatcccagtc tttaattgt	tttggggcca	ggctggggga	ggcttcatga	aagaaccttg	65940
gttttgcgtt aaatccggct	aagaattctt	gctgtactgc	caactgtgt	tcctgtattt	66000
ctatgacact aatccgtctg	ttaaatgtgg	catctctact	aattacagtg	ctgagtagcc	66060
accaggctcc atgcattcagg	tactgcattc	agtgtcctat	tctttgttct	cagttcacct	66120
cagatgatc aattctccta	gcactcccaa	aggttttat	gagagaggac	cagagtggat	66180
tactgtgaa gattcctaag	ctggggggaa	gactgaacat	ccaattccat	tccccccctt	66240
cctccttagc agacataggt	ctagggaaat	tctttgtgag	tggcattata	ctagttggg	66300
aaagggggta gtcaacccgg	aaatgaccgt	ttttttaca	ctccacaact	ttccttgatt	66360
ctgcacttct cctctgagtt	ctgttgtatt	tacaatggag	ctcccatctt	tgaatagctt	66420
ctagtttat ttttatgggg	acagtgtatgc	tagggattt	tctattccat	catctttctt	66480
ctttctgact ttaccctcca	aaagtgtat	cctcaaaaatt	tcaagatttt	tttttgtatg	66540
cctccatgg agctcataca	acaaaacaatg	atacatttgc	aattatgaca	atatgactga	66600
agcataatata gtttctaat	cccagagtcc	tttaattaaa	cacaagaggg	atgattccag	66660
agataacact catcaactgt	attccttagt	ccctctgaga	tttgctttaa	gaggcaggct	66720
ccacctctca catcttgcata	taaatatgc	tgcattttaa	agcttagaga	agcacactaa	66780
acaccccttta atcaaataatc	aaggaaatga	gtatataatca	gaaccagaat	taagtccctg	66840
catttttatg ttcttttattt	ttggttaaag	gggaaatata	ttaatgtatt	aactgaatct	66900
cacaaaatgt ttgcattttt	attattaaaa	atcattttt	tcttttactt	tgcataatcat	66960
tggggggcat gaatcagttt	ctccctctgaa	aaattctcag	tgcctctct	cagaggctgt	67020
gagcaaagaa gtagaaagg	aggtagcta	tgaaaagtgt	taactccatt	tgtaacaata	67080
ctcattacat gcttactata	tgcaaatcat	ttttctaaatc	actttaaatg	agtaaaactat	67140
attattccctc ctaaaaactc	catttctcat	ttttagtgc	ggaacactaa	cgtaagacag	67200
tttagataat ttgacactgt	ggtgggtggta	cgatttgc	ctgatgaagc	tagaatcaa	67260
atcactaatg gggatttagat	tcataaaaatt	gttactttt	gaataactaa	atatttactc	67320
agacattcag tatatttagag	tagggccagt	atcatgtat	ctaatttatt	ttgcaaaaat	67380
tgcaaataatgag cgagatatta	accaagtca	attagaaatgt	tagttggtgg	cagagctgag	67440
cataggttcc agagccaccc	agaacttttgc	tttgcttgggg	agaatctctg	agcacacttg	67500
gatagcttag ggaatttgctg	cttttgggtt	tcactggta	tttgcatctg	atcataggtt	67560
ctagactgct gacaagttgg	ccattagttgg	acaggtat	tgagagttga	ttaagtgtt	67620

gcatctgtca	aggtaaacac	gaaacattct	gttggccctg	agtctcacat	ttcaatctag	67680
aatccatctg	tgtactttac	tttccttgca	aataaatgat	ctagtgaatt	tatataatttt	67740
ttccacttag	ctcctagcac	acatcatttg	tatttcaatc	atttattttt	cttcaccata	67800
ttagaacatt	tccaaatcca	ttaaagaaga	aggaaggcaa	tgaaaaggag	acggagatag	67860
cctcatacat	tctcctttcg	ctgaaagtaa	tatcatttg	aaaacagact	acgggtccca	67920
tcttcttcct	ctctagtttc	caacagcata	aagcttgc	tgtctctgag	gacaagatag	67980
aaaactacat	tagcaaataat	tttccttgc	atctccttat	ttcaagtaaa	tgagggactt	68040
atatttcaa	atatttaag	tcacctgaa	ggttgcttct	taaaaaataa	ctgttttatta	68100
tgcttctact	cttcattta	aaaatttgc	ttgtttccac	cttacaaac	aaaaatggac	68160
aaaatcgtat	tttacccctgc	agtctctag	gacaaattta	acttggggt	actctgtttc	68220
ctaaaaatgc	atatatgtaa	tcttcttattt	ctagtagcag	agtttgaact	caattaattt	68280
ccctaacaac	agactcacta	aaattcgctg	tgatattttt	caaatagtat	tttatggaat	68340
tatcactgtat	gaaataaaaaa	acaaaacaag	aaagacaggg	gaggagccaa	gatggccgaa	68400
taggaagagc	tcgggtctac	agctcccagc	gtgagcaaca	cagaagacag	gtgatttctg	68460
catttccatc	tgaggtactg	ggttcttctc	actagggagt	gccagacagt	gggcacaggt	68520
cagtgggtgc	gcacaccctg	cgcgagcgg	agcagggcga	ggcattgcct	cactcaggaa	68580
gcbcgaagggg	tcaggggagtt	cccttcctta	gtcaaagaaa	ggggtgacag	atggcacctg	68640
gaaaatcagg	tcactccac	cccaatactg	cactttctg	acgggattaa	aaaacggcgt	68700
gccagattat	atcccacacc	tggcttggag	ggtcctacac	ccacggagtc	tcactgattt	68760
ctagcacagc	agtctgagat	caaactgcaa	ggtggcagcg	aggctagggg	aggggcgccc	68820
gccattgccc	aggcttgctt	aggtaaaacaa	agcagccctg	aagctggaaac	tgggtggagc	68880
ccaccacagc	tcaaggaggc	ctgcctgcct	ctgttaggctc	cacctctggg	ggcagggcac	68940
agacaaacaa	aaagagacaga	gtaacctctg	cagacttaaa	tgtccctgtc	tgacagcttt	69000
gaagagagca	atgggtctcc	cagcacacag	ctggagatct	gcctcctcaa	gtgggtccct	69060
gaccctgac	ccctgagcag	cctaactggg	aggcaacccc	cagcgggggc	agactgacac	69120
ctcacacggc	cggttactcc	aacagacctg	cagctgaggg	tcctgtctgt	tagaaggaaa	69180
actaacaac	agaaaaggaca	tccacaccga	aaacccatct	gtacatcacc	atcatcaaag	69240
acccaaagta	gataaaaacca	caaagatagg	gaaaaaaaaacag	agcagaaaaaa	ctggaaacccc	69300
taaaaagcag	agcgcctctc	ctcctccaaa	ggaacgcaat	tcctcaccag	caatgaaaca	69360
aagctggacg	gagaatgact	ttgacgagct	gagagaaggc	ttcagacgat	caaactactc	69420
caagctacgg	gaggacattc	aaaccaaagg	taaagaactt	gaaaactttt	aaaaaaattt	69480
agaagaatgt	ataactagaa	taaccaatag	agagaagtgc	ttaaaggagc	tgatggagct	69540
gaaaaccaag	ctcgagaact	acatgaagaa	tgcagaagcc	tcaggagccg	atgcgatcaa	69600
ctggaaagaaa	gggttatcgc	aatgaaagat	gaattgaatg	aaatgaagt	agaagggaaag	69660
tttagagaaa	aaagaataaaa	aagaaatgag	caaagcctcc	aagaaatatg	ggactatgtg	69720
aaaagaccaa	atctacgtct	gatttggta	cctgaaaatgt	atggggagaa	tggaaaccaag	69780
ttggaaaaca	ctctgcagga	tattatccaa	gagaacttcc	ccaatctagc	aaggcaggcc	69840
aacatccaga	ttcaggaaat	acagagaacg	ccacaaaagac	actcctccag	aagagcaact	69900
ccaagacaca	taattgtcag	attcaccaaa	gttggaaatga	agaaaaaaat	gttaagggca	69960
gccagagaga	aaggctgggt	taccctcaaa	gggaagccca	tcaagactaac	agcggatctc	70020
tcggcagaaa	ctctacaaagc	cagaagatag	tggggccaa	cattcaacat	tcttaaagaa	70080
aagaattttc	aacctcagaat	ttcatatcca	gccaaactaa	gttcataaag	tgaaggagaa	70140
ataaaaactt	ttacagacaaa	gcaaatgctg	agagattttt	tcaccaccag	gcctgcctta	70200
aaagagctcc	tgaaggaagc	gctaaacatg	gaaaggaaca	accggtacca	gccactgaaa	70260
aatcatacca	aatgtaaaag	accatcgaga	ctaggaagaa	actgcataaa	ctgacgagca	70320
aaataaccag	ctaacatcat	aatgacagga	tcaatttcac	acataacaat	attaacttta	70380
aaagtaaatg	gactaaacac	tccaaattaaa	agacacagac	tggcaaattt	gataccaaga	70440
gtcaagaccc	atcagtgtgc	tgtattcagg	agacccatct	cacgtgcaga	gacacacata	70500
ggctcaaaat	aaaaggatgg	aggaagatct	accaagcaaa	tggaaaacaa	aaaaaggcag	70560
gggttgcata	cctgtctct	gataaaacag	actttaaaac	aacaaagatc	aaaagagaca	70620
aagaaggcca	ttacataatg	gtaaagggtat	caattcaaca	agaagagcta	actatcctaa	70680
atatatatgc	acccaataca	ggagcaccca	gattcataaa	gcaagtccctg	agtgacctac	70740
aaagagactt	agactccac	acattaataa	tgggagactt	taacaaacac	cccactgtca	70800
acatttagaca	gatcaacgag	acagaaagtgc	aacaaggata	cccaggaattt	gaactcagct	70860
ctgcaccaag	cgtacctaatt	agacatctac	agaactctcc	accccaaatac	aacagaataat	70920
atattttttt	cagcaccacca	ccacacctac	tccaaaactg	accacataact	tggaaagtaaa	70980
cctctactca	gcaaatgtaa	aagaacagaa	attataacaa	actatctctc	agaccacagt	71040
gcaatcaaacc	tagaactcag	gattaagaat	ctaactcaaa	accactcaac	tatgttagaaa	71100
ctgaacaacc	tgctcctgaa	tgactgctgg	gtacataacg	aaatgaaggc	agaaataaaag	71160
atgttctttg	aaacccaacga	gaacaaagac	acaacatacc	agaatctctg	ggacacattc	71220
aaagcagtgt	gtagagggaa	atttatacgca	ctaaatgcct	acaagagaaa	gcagggaaaga	71280
tccaaaatttgc	acacccttaac	atcacaattta	aaagaacttag	aaaagcaaga	gcaaacacat	71340
tccaaaagctaa	gcagaaggca	agaataact	aaaatcagag	cagaactgaa	ggaaatagag	71400

acacaaaaaa	cccttcaaaa	aattaatgaa	tccaggagct	gttttttga	aaagatcaac	71460
aaaattgata	aaccagtgc	aagactaata	aagaaaataa	gagagaagaa	tcaaataaggc	71520
gtgataaaaa	atgataaagg	ggatattcacc	accgatccca	cagaaataca	aactaccatc	71580
agagaatact	acaaacacct	ctatgcaat	agactagaaa	atctagaaga	aatggataaa	71640
ttcctggaca	catacactct	cccaaactaa	accaggaaga	agttgaatct	ctgactagac	71700
caataacagg	atctgaaaatt	gtggcaataa	tcaatagctt	accaaccaaa	aagagtccag	71760
gaccagatgg	attcacagcc	gaattctacc	agaggtacaa	ggaggagctg	gtaccattcc	71820
ttctgaaaact	attccaaacca	atagaaaaag	agggaatcct	ccctaactca	ttttatgagg	71880
tcagcatcat	cctgataccaa	aagccggca	gagacacaac	caaaaaagga	attttagacc	71940
aatatccctg	atgaacatgg	atgcaaaaat	cctcagtaaa	atactggcaa	accgaatcca	72000
gcagcacatc	aaaaagctt	tccaccatga	tcaagtggc	ttcatccctg	ggatgcaagg	72060
ctggttcaat	atatgcaat	caataaatgt	aatccagcat	ataaacagaa	ccaaagacaa	72120
aaaccacatg	attatctcaa	tagatgcaga	aaaggcctt	gacaaaattc	aacaacactt	72180
catgctaaaa	actctcaata	aattaggtat	tgatggacg	tatttcaaaa	taataagagc	72240
tatctatgac	aaacccacag	ccaatatcat	actgaatggg	caaaaactgg	aagcattccc	72300
tttggaaatct	ggcacaagac	agggatgtcc	tctctcacc	ctcctattca	acatagtgtt	72360
ggaaggttctg	gccagggcaa	ttaggcagga	gaaggaaata	aatggatttc	aatttagaaaa	72420
agaggaagtc	aaattgtccc	tgtttgcaga	cgacatgatt	gtatatctag	aaaaccccat	72480
catctcagcc	ccaaatctcc	ttaagctgat	aagcaacttc	agcaaaagtct	caggatacaa	72540
aatcaatgt	caaaatcaca	agtattctt	tacaccaaca	acagacaaac	agagagccaa	72600
atcatgatgt	aactcccaatt	cacaattgct	tcaaagagaa	taaaatacct	aggaatccaa	72660
cttacaaggg	acgtgaagga	ccttcaag	gagaactaca	aaccactgct	caatgaaata	72720
aaagaggata	caaaccaatg	gaagaacatt	ccatgctcat	ggtaggaag	aatcaatatc	72780
gtgaaaatgg	ccatactgcc	caaggttaatt	tacagattca	atgcccattcc	catcaagcta	72840
ccaatgactt	tcttcacaga	attgcaaaaa	actactttaa	agttcatata	gaaccaaaaa	72900
agagccgc	tcgccaagtc	aatccctaagc	caaaaagaaca	aagctggagg	catcatggta	72960
cctgacttca	atctatacta	caaggctaca	gtaaccaaaa	cagcatggta	ctggtagcaa	73020
aacagagata	tagatgaatg	gaacagaaca	gagccctcag	aaatatcgcc	gcatatctac	73080
aactatctga	tctttgacaa	acctgaggaa	aacaagcaat	ggggaaagga	ttcccttattt	73140
aataaatgg	gcagggaaaa	ctggctagcc	atatgttagaa	agctgaaact	ggatcccttc	73200
cttacacctt	atacaaaaat	caattcaaga	tggattaaag	acttaaacgt	tagacctaaa	73260
accataaaaa	ctctagaaga	aaaccttaggc	tttacctt	aggacatagg	catggcaag	73320
gactttatgt	ctaaaacacc	aaaagcaatg	gcaacaaaag	ccaaaattga	caaatggat	73380
ctagttaaac	taaagagctt	ctgcacagca	aaagaaaacta	ccatcagagt	gaacaggcaa	73440
cctacaaaat	gggagaaaaat	tttcacaacc	tactcatctg	acaaagggct	aatatccaga	73500
atctacaatg	aacttaaaca	aatgtacaag	aaaatcaaac	aaccccatca	aaaagtggc	73560
gaaggacatg	aacagacact	tctcaaaaaga	agacatttat	gcagccaaaa	aacacatgaa	73620
aaaatgctca	ccatcactgg	ccatcagaga	aatgcaatc	aaaaccacaa	tgagatacca	73680
tctcacacca	gttagaatgg	caatcattaa	aaagtcaagga	aaaaacaggt	gctggagagg	73740
atgtggagaa	ataggaacac	ttttacactg	ttgggtggac	tgtaaaactag	ttcaaccatt	73800
gtggaagtca	gtgtgggtat	tcctcaggga	tctagaacta	gaaataccat	ttgaccagc	73860
catccattt	ctgggtatata	acccaaagga	ctataaatca	tgctgtata	aagacacatg	73920
cacaagtatg	tttattgtgg	cattattcac	aatagcaaag	acttggaaacc	aacccaaatg	73980
tccaacaatg	ttagactgga	ttcagaaaat	gtggcacata	tacaccatgg	aataactatgc	74040
agccataaaa	atgtgagat	catgtcctt	gtagggacat	gatgaaattt	ggaaatcatc	74100
attctcagta	aactatcgca	agaacaaaaaa	accaaactact	gcataccctc	actcataggt	74160
gggaattgaa	caatgagaac	actggacaca	ggaaggggaa	catcacactc	tggggagtgt	74220
tgtgggggtgg	gggaggggggg	agggatagca	ttggggagata	aacctaattgc	tagatgacga	74280
gttagtgggt	gcagcacacc	agcatggcac	atgtatacat	atgttacaaa	cctgcatgtt	74340
gtgcacatgt	accctaaaaac	ttaaagtata	ataataataa	agaaatgggt	tttcttctacc	74400
caagaacaca	cacacacaca	cacacaaaaa	aaaaaaaaaa	aaaacaagaa	agacaaacag	74460
tatgctcatc	acacacaaaat	gcaaggtcat	tttttaggcat	tagatggggc	ttaatttaat	74520
ctaaaggaaaa	catgtacttt	gtatcttca	acctgcatac	ctctatttcc	tcttctagcc	74580
ctctacaccc	gttaaccact	tttttatttt	tatctactgt	tatthaattt	tttaatattt	74640
ccccatataa	gtgagatcat	tcagtatttg	tctttctgt	tctggtttat	ttcacttaga	74700
ataatgtcct	ccaggcttgt	acgtgttgta	tcaaataaca	cgatcttcct	tttcaggat	74760
gaataatatt	ctattgtaaa	tatataccac	catttcttta	tctatttgc	tttcattgagg	74820
cacttgggtt	gtttcaatac	cttagctgt	atgaaataata	ctgcaaaaaac	atggaaagtac	74880
agatgcttta	agaggtgggt	aattcacagg	ataaacaagt	ctagagatct	aatgaacaac	74940
atgaggacta	ggggtaataa	aattataccg	tatttggat	tctgtctaaa	tgaatatact	75000
ttagctgatc	ttgcacacag	acacacaaaa	aggtaacta	tatgttaatg	tttgtgttaa	75060
tttgcttac	tatagtaaga	ttttactcta	tgtatcacat	gacattatgt	tgtgaacccct	75120
aaatagatgc	attaaaattt	atttttaaa	agggttcatt	aactgcgggt	caggcaatct	75180

ataaacctt	tacatTTTgg	aggTgagatt	ttcagtatct	gcatttatct	tggattttca	75240
aggggttcat	ttgaacaagc	aaacaaacga	agaaacaagc	ttaagaatct	ttcttacaat	75300
gaattcttcc	taagtagtt	tcttgactgg	tttctttct	ttttctgaaa	cagtctttc	75360
ttattcactt	gtgattgaat	agcctacaaa	attgatctag	acaatttaggc	tcttgaard	75420
gattacggcc	tagtgcattc	tcacctggaa	taaaagcctt	ccttggagc	atcttctatt	75480
tttaattgt	tgagagttgg	ggacttctt	tctcattctg	tatTTTcagt	gtcagTTTT	75540
atggcttgc	tgTTCCAGGT	atggTTgggt	catggtcatg	cagtctattc	acaaggcctt	75600
gtgcttagaa	gaccatccca	agctTggTT	aatgcttga	tctcatggc	ctgaaatttt	75660
acataatttt	tttgtacaag	gattcttgc	gttttatttc	gtattgaaat	aggagagttt	75720
ccttattcctc	ctcgcatg	aagtgcacagg	agtgtggctt	gcttcttgg	tgccccactg	75780
ctcaaatccc	taggggtg	atgcagacag	tcaggtgtac	aggcgtggg	agtgttttg	75840
ggctccaacc	ccacagcagc	gtctagggtt	aagtgtttac	agctcctgaa	gccccaaatgg	75900
gtgtgtgtta	cagtgtgatc	ttttagctt	gctgtcttca	ggcggcttgc	attaatcagc	75960
tcagtttagac	cctctgcatt	accacaaggg	cacagggtt	tctgtattcc	aagttcttgc	76020
ccagtgtatc	ggaaaaatgg	gatcacatgt	ggacatggag	gatgagtgc	aggtttattt	76080
gagtgggtg	ggTggatctc	aatgagatgg	atggggagcc	agaaaagggg	cggagtggg	76140
aattggtctt	cccctggagt	cgggcagtc	atttgccat	gatgctctga	ccacccccag	76200
acgaactccc	ctcggtgtct	gcatcgatcc	accatca	gtctgctggc	atctgctgg	76260
gtctgtgtt	gtgcttctt	gctctgtca	acatccagcc	acttgtgtcc	atgcccacta	76320
aggcctcggg	tttttttgg	cacaggatgg	ggcctgtt	gggcagagt	ggtcttggaa	76380
aatgcaacat	ttgggtgca	aaacaggaga	gcctgttctc	acttaggtcc	ttgggcaag	76440
gtctgtgggt	ggagccctca	ccagggaccc	cacccttctc	tacacatcac	ttcccaaccc	76500
tgctcctgta	tcaatattgg	attctgtata	ctatgtatct	gaggcccag	ttcttaggtt	76560
tacaacttgc	aacttttcc	tatccttgac	ctactccct	gctcgttctt	cctctatcat	76620
gcagaacctt	cccaacat	tgacttagca	agcccagg	ttccttattc	aagaggttt	76680
cacctttat	tctgtctgta	cacagggtt	agaccta	tggtttcc	tctttgtaca	76740
cacatcttag	ctgttagtctt	tttacctctc	tccacttat	agcccttctt	ctctactgca	76800
attgtgtatt	tctggatccc	ataaaagcca	atgagctt	cagtcaaaac	tgtttcagat	76860
ctcacaggca	gttattaagc	cacatgcaac	aaataagtt	ttctgcaat	gaactcctct	76920
agtccaaattt	cctactgtgt	cttttmaat	ctgaaattcc	cacatcctga	tatatttct	76980
ctcattttc	tttttggtaac	cactgaaaact	gtactctt	gaaggccact	gatactctt	77040
ataggaactt	gaaatgagta	acacttaacg	tgagaaagat	ttgaagtgaa	taagagtaaa	77100
gtgaaatatac	tgttagttt	ggagcttagc	tggattttct	gcactttgc	ttaagtgtgt	77160
gtgtgtgtgt	gtttgtgtgt	gcacatacac	acacaaattt	cactcagtga	tagttccaa	77220
agaaccatta	catctggat	ccttgtattt	gctcgtttac	cttctgtaga	cccagaaat	77280
ctgagacagg	tctcagttag	tttagaaaagt	ttattttg	aagggtgagg	acacacctgt	77340
gacacagctt	ccagagggtcc	tgaggacatg	tgttgaaggt	ggtcatggta	cagcttgctt	77400
ttatacattt	tagagagaca	taatctatca	atcaatacat	gtaagattt	ccttggttt	77460
gtcagaaca	actcaaagt	caggtgggg	cagggtggcg	tgtcggttcc	aggctgtagg	77520
taaatttaaa	cattttctgg	ttgacaattt	gtttgtctaa	agacctgaga	tcaatagaaa	77580
agtctgtgtt	atgttaatgc	tggagaggaa	taatgaggca	tgtccgaccc	ccatttctct	77640
taaccgcctg	aacttagtctt	tcaagttaaa	ttttacaagc	cctggctgag	gaggaagtcc	77700
ctttattttat	ttttggttt	cacttcttat	gatattcc	tttagaagt	ttttgttaca	77760
gttgggttcca	taaactttac	ttttttttt	tttttggtag	aaccttattt	tgtcccccagg	77820
ctggagttaca	gtggctgatc	tccgctact	gcaaccttctc	catcctgggc	ttgagcaatt	77880
tttctgcctc	agcctccgaa	gtagctgaga	ttacaggcac	gtgccaccac	acctgactga	77940
tttttggattt	tttagtagag	atggggttt	tccgtgttgg	ccaggctgg	cttgaactcc	78000
tgacccctagg	tgacccacat	gcctcagct	cccaaagtgc	tgcacttaca	ggcatgagcc	78060
atccctgcctg	gccaacacatt	actttttaa	tgaggtttt	cttgactttc	ctttccccaa	78120
aactgtattt	cctttttttt	tctccacaa	agttatttga	attatttttta	tttacttgaa	78180
ataatagctc	tctctgttca	ctagaatgg	aaaaccaaaa	aaagcagaaa	tttttgaaag	78240
gtttgttgc	ttctgcata	atagtgc	atcacacaaa	atattgagta	tcaatagata	78300
tttgcattt	gaatatctat	tttgatgccc	aaaatgacat	gagaagagga	aagataaata	78360
ttatttattt	tatttacaca	tggggaaaaca	cagtgattca	ttcagttgc	aagctttaa	78420
attactgaag	taactctgt	atccagggt	cctaaccagc	tccttaacatc	aagatatgtc	78480
atagcttcta	tttgatattt	tttagatata	tcacatacag	gtactcaagg	gcacaagtca	78540
gtcacatagt	agaacatgac	taagtgtag	gagaatgtgt	ctggcacaat	cctgtttatc	78600
atggtgact	atgtaaagat	gccatgacta	tctgtgtccc	cagaagttcc	ttatcatcat	78660
atatctcaag	gccttccagg	tctatacata	caattttat	aaacagttt	gttttagaaa	78720
ggtgatatgt	attgcagcag	cagtaataa	aggaagttt	taaaatacat	aattttcagt	78780
tcttacctca	gattctcaga	ttgtatgcac	tagtagtgc	acctcggtt	aatttattaa	78840
cctttttgt	cgttgcttt	ctcaatagta	acagcaagac	agagtttga	ttataggcat	78900
tctaaaactt	acagtaaaat	aaatgagggt	actcattttt	agatgagaca	agaatttgaga	78960

agcctgtaat	aaaaatttca	agtatgaaag	aatgggcaaa	aattaaaaac	ttctcgaaag	79020
aaaggaaaat	gtgaattgc	aaagattctg	atgtgaagag	aggtaaaaag	ccaaaaggat	79080
ttgtataacc	tottgttaga	tacaagtgt	gattaaaaga	ttaataaaca	actctgaaaa	79140
tattaatttta	tttctttct	aaacacatta	agtcatgtc	ctaaacatcc	tcaagatact	79200
atatgaaaat	ttattttatt	tcacagttct	ttgttcattt	atttggcata	tgtttattga	79260
gccattctga	ctgtgcattt	tgagctatga	tcttggaaaca	cttgcattag	aatcctacag	79320
ataatcatta	gagatgaaga	ttccagtacc	ccacataaga	atttctgcaa	aaacctgtga	79380
tatgggctaa	taaacaagct	tcccaagaat	tgcttctaca	ctaagttca	ataatcactt	79440
aattctcctg	agaacatgac	atcacgcaag	gctacatgga	gataaatctg	ttagtaaaaa	79500
ttgcaagtca	ttcattctta	catgctctcc	ttttgcatt	ttaagaaaaat	ttagtataaa	79560
aatatgtgtt	gagaagtata	attaaaaaat	gaacttcact	ttgtccctgc	tctcaacacc	79620
attatgtgt	gagaaagtat	atactttata	atgatacaat	gggggctaaa	taataaatta	79680
gagaaatata	gaataactctg	gaaataaaga	agagagaatt	cagtgacacg	tattaggaga	79740
aatggatgag	aaatgctca	caaagaagtg	gcctttaaag	tgggtcatgt	agcatgaata	79800
gaatttagtt	ttaaaaaagaa	caggaagata	gaagagatgt	agctcagtca	tatagataga	79860
tagatagata	gatagataga	tagatagata	gatagataga	tagatgatgg	atagatagat	79920
acaggtgttag	gtatagatat	atctataggt	gtagagaaag	aatatattag	taatgtatgt	79980
cctaggcaaa	tgtatataatg	acattgtaga	cttacaaaat	tatagtaagt	ctaccataat	80040
tacccaaaacc	aaagttatg	gtagtaccca	gaaatagata	actttcgatg	atatatgaag	80100
atgtggagaa	aaactaattc	aaatgaacat	taacatgta	tgtcatttg	tttattttgg	80160
ctttatttcc	atagcttcaa	gagcaaactg	tcagggaaatg	tccacaccaa	atggcagtgc	80220
aatcacagaa	ttcattttac	ttgggctcac	agattgccc	gaactccagt	ctctgcttt	80280
tgtgctgtt	ctgggtgtt	acctcgtcac	cctgcttaggc	aacctgggca	tgataatgtt	80340
aatgagactg	gactctcgcc	ttcacacgac	catgtacttc	ttcctcacta	acttagcctt	80400
tgtggatttg	tgtatataat	caaatgcaac	cccgccagatg	tcgactaata	tcgtatctga	80460
gaagacatt	tcctttgctg	gttgcattac	acagtgcac	atttcatttgc	cccttctact	80520
cactgagttt	tacatgctgg	cagcaatggc	ctatgaccgc	tatgtggcca	tatatgaccc	80580
tctgcgtac	agtgtaaaaa	cgtccaggag	agtttgcac	tgcttggcca	cattttcccta	80640
tgtctatggc	ttctcagatg	gactcttca	ggccatcctg	accttccgccc	tgaccttctg	80700
tagatccat	gtcatcaacc	acttctactg	tgctgaccc	ccgctcatta	agctttcttg	80760
ttctgatact	tatgtcaaaag	agcatgccc	gttcataatct	gctggcttca	acctctccag	80820
ctccctcacc	atctgttgg	tgtccatgc	cttcatttctt	gctccatcc	tccggatcaa	80880
atcagcagag	ggaaggcaca	aggcattctc	cacctgtgg	tcccatatga	tggctgtcac	80940
cctgtttat	gggactctct	tttgcatgt	tataagacca	ccaaacagata	agactgttga	81000
ggaatctaaa	ataatagctg	tcttttacac	cttgcgttgc	ccggtactt	atccatttgc	81060
ctacagtcg	aggaataaaag	atgtgaagca	ggccttgaag	aatgtccctga	gatgaaatat	81120
tgtcatgacc	atgggtatgc	ctttgtttcc	taataaaatct	taaatcgaaa	tctttggctc	81180
acatgtccct	gggttctgt	ggtgatgtt	aatattctt	gtgagtcata	gttgagtgcc	81240
tcagctaaaa	agctcatgt	gggtaaaaat	gagattttc	tagctttgc	tcctccacat	81300
atattccatg	aatcagcagc	atgagctctt	ccttggaggt	tgttacacgt	acagaatcaa	81360
agtctgcacc	tcaggtgcac	tgtatataat	tatgtgttt	atccaaactc	ctagatgatt	81420
gataaggcaca	ctgaattttgc	aggagcactg	ctgtgggtga	aactgtggcat	gccctggaaac	81480
actgttgc	tctttttgtt	tacaacggca	aacaaaataa	atgtgcctcc	agcccaattt	81540
cttgaatgt	ttcttattctt	attctcgcc	gctgcttcag	cagagatgtc	tttaagaaac	81600
ccattttct	geactccaa	aaatccattc	ttctgtactc	cttcctgcac	ttgctgtgg	81660
agaccagaaa	ctagggggca	ttgtgtatggc	ttctgggttt	tataagtgct	tttacatagt	81720
gaagagttggc	aatggaaaaaa	gaggggaaga	gaatgatgtt	acttttctt	aacttgagtt	81780
tatggatcca	gtctctgt	ttacataaaac	ccaccttccc	attctgagcc	tccagttgt	81840
ggaggttact	ttctgtttcg	cctgattttgc	aaaaattttaa	gcaggctgt	ctgagaataa	81900
ggactaaaaat	taaaatggg	aatgaaagaa	ccgagtaaac	aataggtcat	acgatagagt	81960
atccacccct	tttgcataat	tccttctatt	tgtatataat	tttaggttgc	taatgtttaat	82020
tttttggttt	ccttagcact	ggccttgcata	tatgtgttca	aggaagtaag	gattttttcc	82080
acagggaaag	aagacattgc	tctggcatct	tcaagaaact	ttagaaagta	gaaagaattc	82140
ctcttttagac	attatgtctaa	aaagaattat	ttaagggaca	agttaaaaat	atattttttt	82200
aagataagat	ttcactctgt	tgcccagggtt	ggagaacagt	ggcatgtca	tattaataac	82260
ttcctccaa	tccttaggctc	aagtgaccct	cccgccctgag	cctccctagt	agctgggacc	82320
acaggtgtgt	gctaccacac	tcggcttaatt	tttaaattttt	tttgcgtatc	tgatatctgt	82380
ctatattgccc	caggcagggtc	ttaaactcc	gggctcaat	ggctctcc	cctcaggctc	82440
ccaaagtgt	gggattatag	gcataaaatta	ctgtgcctgg	tcaagattct	tctactctcc	82500
aaactttctt	tttattgcct	aaaaataacta	aatattttctt	aaatatcttt	atttggaaaggc	82560
aaattctgccc	attacgccc	tagaaaataa	aaagataaaa	tattttgcgt	atataattaag	82620
ttgtacattt	agtaactgt	tttataaggc	ttctgcattt	tgttgcgtt	gttgcgttgg	82680
tttgcgtt	tttttacaag	aaaaaaagag	ccccaaagct	tctaaatgtca	caacggcaat	82740

gcagttcttg cagacttata tatgtcttat tgctcaggta aggtcggtcc tgggctcaat 90360
aactgaaaat aagcttcagt aatctttacc tctgtttct atggctttt ctttggcttc 90420
aatttcctct agtctccac acaattcatt ttataggatg acataatggt gacgtatgt 90480
atacagaatg ttttagtaact atgaaagcag ttccgtgcta tcttattttt ctagtcaaa 90540
tatttgatt ctatttatt tcattttatt ttatgagaca ggatctcaact gtgtcaccac 90600
ggctaaagt cagtgggtca atcatggcat actggggcca agggagcctc tcaccaac 90660
ctccccagta gctgagacca caggcatgca ccaccacacc tggagatggt ttttttttt 90720
aatatttgtt tttagtagaga tggagtctcc ctatgttgc cagttttgtc ttgaactctt 90780
aggctcaagc ctcagccac caaagtgcata ggagtaaagg catgagccac attgcctgac 90840
cttgcattt aactggagcc tgaagaagcc aaacttattc ccaaatgaac atttccacaa 90900
gtaactcaga tatcagtaac atttttagaa atttacttt cacttttattt gaatcaatgt 90960
aattttttt ttagttgggaa ggatggggaa atcttggAAC tccgaatttc acccaaataat 91020
ttaccatgtt taccatgtt ccataaaatga cacctcttagt tttctgttca acaatataatt 91080
caaacagsct actaaatgac aaataatgtt acttgaaaaaa tatttttcattg taataaaarat 91140
aaacatatgg atatagatat cctgtgaatc aaaatttacc aacatgraaa atatgttgg 91200
actttgttat ctatgttata ttttagtaatt ccaatacagt gcaagcaag ttaacattat 91260
tgaagcagt aaacatttgg agttttattt tggttatatt tcttgcataa agaaacatag 91320
actttttttt aactcactgt gtctacactg attcctaattt acccagaaaat ttcagactga 91380
acaccaatgt catggaaaaaa ccccttgcata gaaggacatc tggtttaatc catcattgag 91440
gcttaattt tttctgtgt gttatttctc acacagagaa caatgtatgt atagcttatt 91500
caaaactgtat tcagcactt ttctaatgtt ttaatgtgaa ttacttggca atcttatattg 91560
gaatacaagt tagaaaataa gagtcatgaa agataaaattt tgctacttt tacacaaata 91620
ttaaataagt gtcactttt tctaagcaca gggcacactg cgcattgtaa acatcaaata 91680
agttatcaaa ggcatttc acaggcatta aacatcacct ttgatcatt tcagcaaccc 91740
agtttaacca ttggcaataa aataacattt tgcttaagttt cacattcacc aggcttcca 91800
taaaatatta ctcccagca cctggattaa ataccatgaa ctaactctac tacaagaatt 91860
ttgaggtcca aattttttt cccagtagaa gtttaataa agaactagaa atgatctca 91920
gatgacgtgt tattaaacat ttcacaatca tttagcttattt tattttcattt agactaaagg 91980
ctttagctct gatgcagaaa atttgaatat atcatgggag ttagtcaaaat tattttaaat 92040
ttaccaatag tatttgcatt atgcattgtt ttagtaatct cagttgagca tatgcccccc 92100
caacatataat atacacatac atgaatataat atagagggtgc ttgtatgtt cactgaggcc 92160
ctctagaagt gaaacttggaa tgtttaataa tgactaagtt cataatctcc tgcccaggaa 92220
attataaattt aaaaacattt cagatctcaa agagctttt tttactgaaa tatttgagaaa 92280
gacattccc aagtacttta gcctttaaaa tatgatttcat gcaagcaact aaatttagaga 92340
gttcttggga agagggctgt ggataacaag agaaaagtga gaagagaaga aactaaggat 92400
gtgtgatcaa agaacaggtt aacatgtaaa aaattcaaaa gagaaaaaaa atgcttgc 92460
gttataatgag agtattttaa atatttggac gttgtctcat ggattacatc tatttatcta 92520
tctatctata totatctata tatctatcta tctgtctatc tatctctcca tctatctata 92580
atctatctgt ctgtcttata gtttgcattt tgccttacaa ccaataaaaca cttaaataagc 92640
aaccatgatt atttgcattt tattttttt tattttttttt tttttttttt tttttttttt 92700
cacaatacag attgttttag agttagaatgg attcacgtcc acctgcaattt gatacgcaga 92760
gctctgatatacatttctgt tttttttcc tcagagttt acattttttt ctcctgtttc 92820
tattctctcc ccatttatgg agtgcctctt cagtgctcag tattttttttt aagcattttt 92880
tttcctataa tcaatgtat ttctctacta gtggcacggc tgaatgaagc tggtttccca 92940
aattgtggag attatgtctt acttccttga attttgcattt gttttttttt tttttttttt 93000
acagcattcc ctgctgctca gtgtgaaattt agggattttt gccttaaagag aaaatttgat 93060
agtgccttg ggtcttggaa taaggaaaggt tgggacatgc tgaagcctt tgcccacttt 93120
ctggagctttt actttacaca tgcctatataa acaggatctg tgatgttctg caatagagca 93180
acctatttttta cttttccccca cttaaaacccca gtatttccaa actttttttt ccccataggc 93240
ctattttgtt gtttaatattt cagtggttct tctgaacaaa gttttgaaaa ttctaaccta 93300
aaaactcaga cgatagcaag attaacatgt ttacaacccc ttttccttac tcctaatgg 93360
acttgctcta catcagttga agggcttttcaaaaatgtgaa atgaaacaga agcgggtgtga 93420
tttcttgcattt gacctaaagt gcctatgtca atggggaggag atactattat ccaaatatga 93480
gttttctctt taagctgttag agggtaggtt gctgcattata tttaaaggaa agagcaggac 93540
atagaaaattt aaatttataatgtt atcttttttta tccttttttca tttttttttt tcattttttt 93600
attatttgattt aagtataataa tattttttttt tttttttttt tttttttttt tttttttttt 93660
aaatgtaaag attagataaa tatattttatg ccatttttttccca ataaagatttta tattttactaa 93720
ggtgaatattt ttccaaaatg atgcatttttctt ctggagagtg agggaaaggcc acttttactc 93780
tttacataaa tctttttttt ctaatttttgcattt cattttccctc tcaaaaggcca aagttgtaga 93840
tttaagacaa gaaatcattt caattttccaa gtaacaggga acgtcagttc taattttatca 93900
ccatcacttg ggtgccatc acacgcacag gcacacacac tcaaaaactg cgtgtcaag 93960
acaatccaa agatgtgctt ctttgcattt ttcaggaggc atgttctactc ttcttggat 94020
tqgttggaga cttaagttt ttttgcattt cttaccgctt attaaatgtc ttattttgcca 94080

caatagaaaa	gtcatccctg	catctatgtg	aaatatgtt	aatttcttt	gaaagttgcc	94140
taagacaatt	ttctgcatgt	actgttcata	ctagctaaaa	ctgctcccca	ctcttattcc	94200
tctaggaaat	tcctagttat	tttcaagcc	ccagttagat	tattgtcctt	tgatgcttac	94260
cctgattcct	gaaacaattt	gttattttgt	ttgtatTTTT	attattgaac	taatcatatt	94320
taacttaat	tttcatgctc	cttaccatga	aaatcaacca	gctcttcaa	ggcaagcact	94380
gtgatcagtt	gtcttcaatt	ccccagcaaa	gcaacttgca	tgcatggagt	gttcagtgct	94440
gtttagtcac	aaatgttaacc	atattacaat	ggttaaatca	tttagcatcc	tgaaagcatc	94500
acagagtcaa	agtagctaac	ttgtgtgaac	ccttaattca	atttccggga	agaggttaga	94560
cactgaagta	tgttagttgaa	gacactggtg	ttgcaaggaa	ttgtatccca	ctagctcaag	94620
cctacagact	tctgtttgt	gccctggatt	aagataaaatt	taaaaatcga	atgaattgcc	94680
aacactatct	ttggtagatt	tcccataaaa	acaatgaatt	ctgatattta	aatgtaaag	94740
ctcttaagac	cttacaacac	aaggccctg	ctaacataaa	gcaaaattct	attggcctga	94800
gtgacttctg	agtctacact	gggaaccact	ctatgttca	cttcagtgg	ctcttaaacc	94860
caaccactt	cactcattct	tattagctaa	agggctccca	tagacattta	agtttgatc	94920
tcttatctcc	tttagtgtt	tacttataaa	ataacacaag	gcaacattgg	gaatttataa	94980
taaaaaattt	ttgtatagaa	tcccagcacc	ataatttccc	attgtctcat	ttcattacaa	95040
ttgatattat	ctgaatttaat	attatctgaa	agtataatct	gttcattata	atttgtattt	95100
tttgaactct	aggaataatt	ctatctct	tgattactt	tcatttttatt	attctggaca	95160
ttataatatt	ataaaagaagt	ttgaactaat	gctgctagaa	acctgaaaca	aatttgtgtt	95220
ccactagcg	aacaaatttg	tcacattctg	ttatatgctg	taaaccctt	aatgcattc	95280
aggattaaat	taaaacatcaa	atggaccagt	agataaaatc	tatgaaataa	cattggcaa	95340
cttcttgc	tttactccta	caaactcctt	cgactgaaaa	ttccttcatt	agtaattcat	95400
ataggccctt	gttattccaa	atttagaaac	cttacatttt	gtttattcac	aaagaaatga	95460
attgctttt	attttgttgg	tatagcagta	agaaaaaaa	agtgaagaaa	aaaaaccagg	95520
aatttccttc	aaacccctat	gattctggac	tctctaatac	agcacatagc	cagtgaaatt	95580
ctcaaagggt	gagacactaa	agcaggggtc	ccctaacc	caggccatga	actggatgg	95640
tccatgtcct	gttagcaact	gggcacaca	gcaggaggca	agccaccagg	aagtgagcat	95700
tcctgcctga	gctccgcctc	ccgtcagatt	agtcttggca	ttagattctc	atgggagcac	95760
ccaccctatt	gtgaactgca	tgcgcattgc	agctatctag	gttgtgtgct	ccttacgaga	95820
atctaattct	tgattatccc	caacaacccc	acccttcaac	cccatccat	ggaaaaattt	95880
tcttccacaa	aactggctgt	tggtgctaaa	aagggtgggg	actgttgtt	taaaggaagc	95940
ccatgattgt	atagtcaact	gaaaaaggca	tttagtggct	taatccattt	ttcgacaaac	96000
tttaggacac	tattccctca	ttccaaaaca	cgttggtaac	acagaaacca	ttctttttag	96060
gaatcatctg	catatgatac	accagttaa	ttggggcctt	tgcttttttta	agaccaccag	96120
gccctatatt	cttgagtcct	gcctgactgg	ctgtgtgagg	gagaacaaga	gagaaagaaa	96180
aagagagaga	gcaaaataaga	gaaagagaaa	aaaggttaaca	tgtatggac	ttctcattca	96240
ccttgattta	ttttcatcca	ttaagtgtaa	aatgcttgg	acctaagaaa	atgtttagag	96300
gaaattctac	tttggtgacg	gaatttattc	tcttgggatt	aaaggatctt	ccagagcttc	96360
agcccatctt	ctttgtactg	ttcctgctaa	tctacctgat	cactgtcggg	gggaaccttg	96420
ggatgttgg	gttgcattgg	atagattcac	gcctccacac	ccccatgtat	ttctttcttg	96480
ctagttgtc	ctgcttggat	ttgttattact	ccactaatgt	gactcccaag	atgttggtga	96540
acttcttctc	agacaagaaa	gccatttcct	atgctgttgc	tttagtccag	tgcttatttt	96600
tcattgtgt	ggtgattact	gaatattata	tgctagctgt	aatggcctat	gataggatgt	96660
tggccatctg	taaccctttg	ctttacagca	gcaagatgtc	caaagggtctc	tgtattcgcc	96720
tgattgttgg	tccatatgtc	tatgggtttc	ttagtgact	gatggaaacc	atgtggacat	96780
accacattgac	cttctgtg	tccaatatca	ttaatcaatt	ctactgtgct	gaccaccccc	96840
tcatccgact	ttcctgtct	gacactttca	ttaaggaaac	atccatgttt	gttggtagcat	96900
gatTTAACCT	ctccagctcc	ctcatcataaa	tcctcatctc	ctacatctt	attctcattt	96960
ccatcctgag	gatgcgttct	gctgaaagta	ggcacaacc	gttctccacc	tgcgggtccc	97020
acctgggtgg	agtgcgtgt	ttttatggaa	ccctgttctg	catgtacgtt	agaccccca	97080
cggacagggtc	agtggacacag	tccaaagtca	ttgctgtttt	ctacactttt	gtaagcccta	97140
tgttgaacccc	catcatctat	agtttggaga	acaaggatgt	gaaacaagct	ttttggaaac	97200
tgatcagaag	aaacgtgctt	ttgaagtaaa	atcagtgtat	ctttattttgt	caaataaaaa	97260
aatctttcta	tttatgagaa	ctgttattaa	cttttagtagc	tttacagttaa	agtcaatttt	97320
agatccccaa	tgaaaaacta	taatctaaca	acaacacata	gacatagata	gagaaaattt	97380
atggttttac	atgttacata	tgagtgaaat	atgtgtgtgt	gaagggttagg	gaggtgtata	97440
tgcatgtatg	aggaataattt	attcttttat	tcataacatg	catacatat	catgtatgaa	97500
tttacatgtg	taacaattta	catgtaaatt	tacatgtaaa	ttacatttat	aacaacat	97560
tttttacatt	ataaaaaattt	atacattgtt	gtaataaaac	attaggtctc	tatggaaagg	97620
taaaaagtaa	gagattaaaa	tatttttttt	tctttccac	accctgattt	tttattctca	97680
cttccttagag	ttacaacat	taacagtttc	tttagttttt	ttttttttat	ttcttcgtat	97740
ctacaaacaa	atgtgtgggtt	atatgtacac	acatataatac	acgattgtta	aaacacaag	97800
tgactgtcat	ggttgtttttt	tagtggttta	caattctgt	gcaattacta	ccaagaaact	97860

tcagtctcat ggtttagagg aatgtcacaa ctgctggta gacttgctgt ccctcttaggc 97920
 tggttctcta ggagagaaaaa cttacgataa cacctctgtc tccatgcctg acacccactt 97980
 atgtctctcc acttgtgaga gtcactttca ataagacacc tttctcttc tgccacgaca 98040
 ggacatgagg ccagccagc agttatTTT caaatattaa ctctggaaagc agacaaagaa 98100
 tggctcctat aagggtataa gactggctgc aaataactgt gacttctgca gaacttataa 98160
 attcaattct acacttctgg ggacttagagt cttcatgct ttctctcacc tctgctttt 98220
 taatctatTT ttcaggatag atcatatatt tttatttagta gcatcattca aattttactc 98280
 acagtggta tttgtcacag ttagttatta tatacctaatt tttagaatga catatgtcga 98340
 cttgtttgt agtatTTT tagctaaagga agcaaaggac aagttgatTT ttaaggatg 98400
 tgacaaatct ggtgactttg cagttgataa cataagcaaa attgatgaat gtgtgctgag 98460
 tatctacaat aagtcacact tctcctaaca ccaaggactt ttcaaggagc tcgacaatgg 98520
 catagttta caggggtcat gaaactgtaa acagaaagat caatacaatt tgtaaatgc 98580
 tgtaagacta tggaggtgt aacgaaagag aggaaagaaa aaattatctc tgggcaaatt 98640
 gaagctagcc ctgtgattgt ccaagactag aacttggTT tgctttatgc accttctca 98700
 gcccctccat agctcctagt atgacgtaat gtcaggtggg gtctgagcaa aataattaag 98760
 aaccaatttT tgatcaaata ttagatgtt agttcaatg aaaaagaaaa aaccataaaa 98820
 ctataacaag tcacaactca aaaagacatg atatTTT tcagttgtat ttcaaagctt 98880
 tctttgtacc ttcgagttaa tctctgctt aTatttctac agaaattttc agacttattc 98940
 attttatattt attgcatata attatTTT accttggact actacatgtt cctatttaaa 99000
 taccttgatt ttgaaaagtca gattgcccac aaattctgtt aaaaacacaca tataaataaa 99060
 aaagctattt ttactaattt tttacattt attaacaatt gttgaattcc gcttagacac 99120
 ataccttagat gtattcctgt ttaaatacac atatgtttt agttatgtga aaccagctt 99180
 ttaattatTTT aaaaatttttT gatctcaggg cccatgctt tattctgtaca gcctagctt 99240
 gcaccgaaga aagggtcaat gactgcagca taaaagagta cacatcatac acaattttat 99300
 taaaagttaa tgaatgtcca ctttcaaagt aaacaaagtg cctcaatttt agaatttata 99360
 agaaaaagtgt ctttgaacag accagcagat gactctgagg gaatatgtta tttagtataa 99420
 gccaacttac ttttcagaaa tacaagtaac cataatgtca ttgatgtt aTactgactt 99480
 tagcaataca aaaccaatag ttacgcccattt atacacaaaa gttcaagata gaaaggaatt 99540
 acaccaagtg ctttcatgaa ttgactggct cgctggaaatc acgtctatca gcctgttaag 99600
 actatctaag tcacatcat attgtatttT acataaaaata aatatgtgt gatgaaaatt 99660
 tttggattttt atatgtttt tttagtaggccc ctggattttt gctaaaaagt aaagaataac 99720
 tttcttatttT atgacatttT aatccattt ttaaagtctg tcataagaga ttcaagaag 99780
 agtctccat ttcatgtctc ttttcttca aagttatttT tgctgaattc ttattttgtt 99840
 ggacaaatttT ggtatataaa tgaagaaata atcagaaaaa ataaagaata ttactgtac 99900
 ttttggaaaag aacaataaaat atgatataatttTtcttcc cacctatcca aaagatcatc 99960
 actctgatTT tcctttcagg aaaattttTt catgtgtata aaatactaaa tggatagg 100020
 agtagtgatg ttatgagcct aatgtcttta ctcataatac agatgtttt gggacaatg 100080
 aataagtgaa aataattttTt acactcacac tcacacatc acacaaaaat aagacagagt 100140
 aataacttttTt attatgtttTt aggtatattt ttaatgttaa ttgtttact taatctgccc 100200
 atcaacacta tgcattcagt actataatct ccacttttttTt gaaaagacaa cagatttaca 100260
 agttgttaac ctaaaaagcac ataggttagat tagaaggggg gctcaaatat gcctggctt 100320
 aaagtttttTt tggggTTTt tgttactatt atacttttgc cactatactt cattgaaggc 100380
 aaaatcatac aatgaaaaat ataaattttt tctatcttcc agaagtatttTt ctgatcatgt 100440
 atttgacagc tgcattcacta ctgaaggata taatgcacta taaaaatttTt ttttataat 100500
 gtatagataa attaactttTt ttccataaga caaatttttTt cagtcagatc ctaactccat 100560
 catttatttTt ctaaatgat agttatgtgc atgaactctg gaatctgatt tactgagcaa 100620
 gttatacaac ctctctcgcc ctcagtggcc tcttctataa agtgaagaca tagtgatTTT 100680
 tcataaggat gctgtctca cagaagatTTTt taattaacac atgaaacaca cttaggacag 100740
 tactggatTTTt atagttatttTt ctcataaaaaa ggtacctgtt ttaatttatttTt actggaaagt 100800
 tactttacTTTt atcttagact tagtttctt tcataacatttTt ctatgttttTt ttttataaca 100860
 ttttatttTt tttggatata attaatgtct gtttctctt taatatttTt agtgacatga 100920
 ggtcagaaaac aatgtatTTTt tgagagcaat tttttctt ttttgcctt ggtatgtatg 100980
 atgtgtatccc taaaaatgt aTatgtatTTTt attgtatgggt agatgacaga tagatgatag 101040
 atagatagat agatgaaaaa aaagaagtaa aaaaaataat ggagattata atattatTTT 101100
 tgtaaggtaa tgggatTTTt ctgcacaaaca aaacactgtat ggttaacttta atgtgtatTTT 101160
 ttttatttTt catataatcagt ctgttttgc ttccttcaac agaatttcatc tggaaatgtg 101220
 cagaataaaat caaaccagag agaagaaatttTt tatcttctt aggttggccca gttggccggc 101280
 gttgcatttTt gtcatttttTt tcttggTTTt attaatttTt atttttctt ttttggaaaaa 101340
 cttggcccta aatctgcaca tcacatcctc acactgtctc cttgccaatc tctcaatcct 101400
 acaaatttctg gttaaacttttTt ttttggcaat gttcaatttTt atagccaaat tataatgaaaca 101460
 gcataacatg gtcacatcc cgttaacccca gtttttgc tctgtggccca tctgtggccaa 101520
 tcacatccac atgggtcatct gccttccatg tagtcacaat attgggtcca aagtgcctca 101580
 cgtttatatg gctgtccttc tctatgttttTt catttcatgt aaacttatttTt aacatggatgt 101640

acttcttaaa gactacatgt gtggctatca tcaaagtgc ggtagaaata gggtagcaa 105480
 aggccattct gaaaagatct caggtgaaac tgagaaacaa tgtactggaa actagagtaa 105540
 aggccatctt ttttatacag tggcaaataa attggcagaa ttgtgtccat gtagcagggc 105600
 tctgtggaat acagaactta agagtggta gctagactat ctggaagaag aaatatctgc 105660
 gaagcaaaaa attaaaggtg gtgcacagct tattttgggt gcttacagta aaatgagaga 105720
 agaaaagaaaa aaaagataga atttgtaatg aaaaggaagc agcatgaagc gatttagaaa 105780
 actcttagcc tggccatgta aataataaaa aagcatgttt gggacagaat actaaggaag 105840
 tgttcatgta ggatacagta aattccttcaa aagtttttag cctgttaact tcaagaagga 105900
 gaaaaattgt taagtacaaa gagttctgag ttccactca aagaaccaat caatatgtca 105960
 gtatgttag cttccctgtt ctttgttctc cattttaaag ttaacactcc tcgttcttta 106020
 catctccctg ccccttagttt cagtaaacaac cccctcccttgcctctatca cctgttctgt 106080
 ccttagtcat cctcagtca ctcgtctgtc tttagttatc tagtcatcta cactgttaacc 106140
 gtcctcccg ccgaaactac tcaccctgccc actctggctc ataccctgc tctctttaaa 106200
 atagccagtc agaattagct tagactgtgt ggtccaaccc cagccaatag gggaaagaca 106260
 gcagtagggc ctagctgott taggataagc acctcttccc cttccctgttgcgtgtc 106320
 tcgccatcgc tccatccttgc agatgcaccc ttctatagaa gtaaattgccc ttggtagagaa 106380
 aacttttgc ttagtgc tagtttacttgc tagcaccgaa catttacttc caacaatcat 106440
 ttgctacaga gatgaacata aatagaagga atccaggtt gatttatcaa gacaatggga 106500
 gnatgaccct gaaagcattt caaagatctt gtggaaaagc tgggattttt agggcaaaga 106560
 atccagagag atacataaat atatgttttag ttttaactg gaagtgcgtt gggatgtact 106620
 gtatgttagt caatcatata ttttaacatgt gtggaaaactg aaatagggtt actaggctaa 106680
 aactaatggc ctagcttgc aaaacaggat ttctcgctc aaaaaaaaaa aaaaaaaaaa 106740
 aaacagggtt tctcagttaga actaaactgc cttcccttgg ttttattttctt agttaatgt 106800
 cgtaataaat ttaccagata ttgcttccgt gtacctggaa agcatagaca ttttgattca 106860
 ttgtaatcaa tggtaatca tggaaattaga tggaaataaca gcaagtgaag catgttagt 106920
 catgttagt gctagagaga ggagattcat agtcaaagct gtcttcttctc tcagtgagc 106980
 tcactcaaca gcttttggaa tttgagaaaaa ttgaattttc tttctaaggc tcagtcctt 107040
 caatatataa tgggtctaattt aatattttt tcataattttt aatatttaca taaaattata 107100
 cataaaaaagt tctgtacatt gaatttataa aaactttta ttttaagtat acttaattac 107160
 cccaaattttt aacatctgta tgggtacat catggcttag gaaaataaaag aataataaaa 107220
 attatcctaa tcatttgcac tctagatttc ttcaattata ttcatattact cttggacttt 107280
 tctacaaaca aagcaacaca aaataaaacaa agaaaaagaa ggggagcaaa aataccctat 107340
 gactttttt ttttttcaa atgaatgtat gtcaccgtt atgataccgt gtggtaga 107400
 atgttttgc tgcatttagt gccaagtggaa aagagtgtt aaggtgttcc ttttttagt 107460
 cagtttttc tggcaagttaa tggagggagt gaggtaga 107520
 aagcaataag taaagtctt accaagagat gtgaaaaaaga ggacacagaa gagatgtt 107580
 gccattttaa atccatgtt caacatcaact gagctgtctt cagggacat ttaattggca 107640
 gaaatgactg cttctttgtt attttttttt aatgcaacag gggaaaggag tcaactttat 107700
 aaacacaaggc aaacctttcc ttattttataa tctctgaagg cccattgggtt ttcctgctt 107760
 tttataggtt gggctgtcta tttgtgtgt tttgttagga actagacaaa gtgttctat 107820
 ttatatttag aagccaaata cttattcagt agaaatgtt catgatttataa aatagttaaa 107880
 taacttcata ggagttatggg tggttatataa ggtatcaatg cacacaatct gagaattttgt 107940
 atttggctta actactgaac aaatggact aagaaagtaa aatttcttgc aatggctta 108000
 tggcctcac atgttgc tagtcttacat agattataac ttttaatcct caaaagat 108060
 tccttcaaca acatttggaa gtgagaaaaac taaagctaa aaaagttatg taagctgtct 108120
 acaattgttag agctaacatg cagcatagtt gtaattttt aagtcaaata aggcaagttc 108180
 caaaaactttt gagttgggtt tggagatga ataccagtac taaggatgtt gggaaacattt 108240
 tagcacaaaa taattttaat gttttggaaa gtacaaaatgt acatggcaga ctatttctaa 108300
 ggaaggtggg atgtaaaggaa aggacatgaa tatatagaag agtgcacat gatgttccag 108360
 tggctccag aaaaaagat gaagaaaaga ttgcaaaagca tctaactacc caatcttcat 108420
 gaaaacatgtt gcatgcacac tgaaataattt aaaaactaca tgggtatgc tgcatgc 108480
 tacacatgtg tggatttat atgtctatgt gtatacatac aatatacata gtcaggaga 108540
 ctttatatac atagtcaggaa gaagagagag agaggtaaa tcaccatcct tctaaagata 108600
 atagttccta ttacatgaa acgccttttag ttcgtccaaag caccttaatg tggacttag 108660
 gcaaaagggtt agctagatgtt gaaatttgc ttggagacc cccaaaagggtt aagaacttag 108720
 aataaaaaata tagaaaaaa tacatgtttt gaggtctttt tagaagagaa ggaaggagag 108780
 gtgagaaaaa ggagggggagt agaattaaga gaaaccaggag agtgcacatg aatttgggtt 108840
 agtgcgttac aataaccggt gacaaatcctt ctatgatgtt gcaacaggta gaaaatttga 108900
 gaaatataat agatttaaatg ttttaattttt aatttttttctt ctttttgc acaaataatct 108960
 atcttctact tgcctccac agagttgc tattatctgt gtcgtgtgtt ttgcataactt 109020
 gtaaacatca cagatttcta aatggcgtgg aatagatagg agttaaaga tgattttgaa 109080
 aggtctcaag ttgctgacat caggatttac taagcattat ctttacattt tataagtc 109140
 accatagtcc ttgaatataa tctctcacag attaattttaa ttctactaa taacaatcac 109200

ttcatatgtt ttcatcttat atagcagtca agtttacaat aatcttatta agcaccaacc 109260
 ttaagagtat ttatattatg tagataacgg attaaattcc tatgagaaca aaacaaaata 109320
 acagttaaac aaattagaag tttatttctc atttatgtaa aattctgagc tagtatgtca 109380
 tatcagctct gccaagggtt taggagctcc aactaattta ctcttctttc tctgcttct 109440
 ataaagagtt actctcatct atatgattca ggatggctca gcactatatc cacattacag 109500
 ccagtgagag gcagaaaaga agaagggaag ggc当地ctg ttttatcaat ggatggataa 109560
 aagtagctt attttctttt attacattct ctggccaga agctaccctt ttgc当地tagc 109620
 aaatctaatt ttagaatttt aaaaaaaagga aattctggtc attttctttt tgcttctt 109680
 ggcttgggtt tttaaaattt atttaactta ttatattta agagctgtcc cataattatc 109740
 agttctggct ctgaggagca aatacttaac ctggtaagg gaaaggtaact agcttcttagt 109800
 gcaggctcag gctctgtcaa atctgtctgt tgc当地ggatgat cacttagcca 109860
 tctaattact tcgctatgca tgctcagtagt gc当地aaat agacagacat gttggacaaa 109920
 gtcatttcca tgctctaatt caggctgtc gctccaaaaa attacctact ctgaagttt 109980
 tgaaattaac tagaaggcatg ttagtagaca ttttggaaaat ttttctgtaa aattttcaat 110040
 tggatcctc ctcaaataatt tatctttat tttccccaa aagtagcagc tagtctaaat 110100
 tagcatattc tggatcctc ataaagaatg agctaaaatg tatttttagt ttttactac 110160
 ctatccttat aacaggaacc taagggattt tatctgacat ggc当地ggaa aattaactgt 110220
 aatcagaaaaa agaatgtcag gtaccagggtt ctcaacttac agttaagtt gataacttat 110280
 agctaagttt tctaaacaat taatttacaa aagtgc当地g tatttctttt tttttagca 110340
 ctagtggcaa aatctgagca ataaagatgt gagataaaaaa agcaaaaatgt aaaaaaaaaat 110400
 agacacagta gaaatgaata tgagtaagaa attaagtgc当地 ataaaccgaa aagattagac 110460
 cctgttccaa ttaaatctact atctcttagt tctc当地t gaca caaaaacaaaat ttagaaaaac 110520
 tgcagaggag caatgc当地aa caatgtttca agaaaaatata aaaaatgtttt ccctttagaa 110580
 aatattctact aatccaaacat tctc当地ttcc cactgaagga aattatgaga agaaactgt 110640
 cattggtgac tgagttcatt tctc当地ggac tggccaaatca cggggaaatca cagatttcc 110700
 tcttacacgct gtttctcacc atttacatgg tcaacgggac agaaaaatctt ggc当地gattt 110760
 ccctcatcca ggccaacgccc cggctccac acgccc当地t accttttccct gagcaactta 110820
 tccttggatc atctgtcatt ctcttccaaat gtgactccaa ggtatgctggatc gattttccctt 110880
 tcagagaaga aaagcatttc ctatcctgccc tcttactgtc agtgc当地tccctt ttttaccc 110940
 ttggtccacg ttgagctcta catcctggct gtgatggcct ttgaccggta catggccatc 111000
 tgcaaccctc tgctttatgg cagcagaatg ctgatggaga ctatgtggac ctacaaccta 111060
 gtgctttatg tggatggagc actcactgtc ttctactgtc tgacccacc actgattaag 111120
 gccttctgtg gccccagtgaa attaattcac ttttataccca ctgatggcct ttgatggc当地t 111180
 ctggcttggatc ctgacacccca caacaaggag gtgtcaatgt ttgttggc当地 tggatggc当地t 111240
 ttcacttattc ctctccttat catcctcatt tccatctct acatatttcc tgccaccctt 111300
 aggatctgct ctacagaagg caggcacaat gcttttcttca cctgtggctc cc当地tgc当地 111360
 gccgttacta ttttcttattc agctcttttcc tccatgtatc tcagacgtcc atcagaagag 111420
 tccatggagc agggaaaaat ggtagctgtt ttttataccca ctgtaatccc catgttgaat 111480
 cccatgatct acgtctgag gaacaaagat gtgaaaggagg cattatgcaa agaactgttc 111540
 aaaagaaaaat tggatggatc ataaacattt ctactgattt ttgttggatc gtc当地ttt 111600
 ttaccctatg attttccata gagcatagtc tgatacaat atatccaaa ttatatattt 111660
 tccttagaattt gttttttttt ttttatttttcaat ttttataccca ttttataccca ttttataccca 111720
 tcaattaaaca ggaatttttcaat ttttataccca ttttataccca ttttataccca ttttataccca 111780
 gctctatagt gaagaatccaa tggatggatc ttttataccca ttttataccca ttttataccca 111840
 aaatccttaa aaccagaccc ttttataccca ttttataccca ttttataccca ttttataccca 111900
 ctgcttccat aacatatttcaat ttttataccca ttttataccca ttttataccca ttttataccca 111960
 caaaaaaaaac acagcatctc actccatcac ccaggatggatc cctgatggc当地 gccc当地t 112020
 cacagcttac tgcaatcttcaat ttttataccca ttttataccca ttttataccca ttttataccca 112080
 cctctgccat agtagttggg actacagggtt tacaccatga tacctgacta attttctgtt 112140
 ttttttttgc当地 agagacgggatc ttttataccca ttttataccca ttttataccca ttttataccca 112200
 ccagtaatctt acctgc当地t gcttccaaaat ttttataccca ttttataccca ttttataccca 112260
 cccagcttac aataagctt aaaaaaaaatca ttttataccca ttttataccca ttttataccca 112320
 tggatggatc aaagaaaatca aacatccatc ttaccatcat tcaacctaaat atatggccct 112380
 catcccttgc当地 acactcagct gtc当地tccatc ttttataccca ttttataccca ttttataccca 112440
 ccagccaaaaa acacccaaac acttggatcc acagtggagc acactttagt ttccacagtc 112500
 catacctggatc ttatttcttcaat gcttccaaaatca ttttataccca ttttataccca ttttataccca 112560
 ttgatggatc cttaaaatca ttttataccca ttttataccca ttttataccca ttttataccca 112620
 gcatgtctaa aaccacatca accaaaaagg gatccaaaggc taacagaccaa ttttataccca 112680
 cacagatttgc当地t gatggatca ctgtcttccatc ttttataccca ttttataccca ttttataccca 112740
 ggcctatagt tcccttagat gtaaccttaga acttaagata taatcacata caaggaaagcg 112800
 agagactaat gatggatgtt gtc当地tccatc ttttataccca ttttataccca ttttataccca 112860
 ttccgattat tgaagggatc caacctcttagt agtaggacag agtcaacaac agcaacacaaca 112920
 aggatgaaac tagtccgtt agacttggatc ttttataccca ttttataccca ttttataccca 112980

aatttgcataataat gctcatgtta agtcagcatt tatatttata ctgcgttgc 113040
gggctatcaa atctcaagct aaataatgaa gaacataaaag aggcccaggt tagaagaagc 113100
ttcatgacac tgaataagc aaacataaaac ataagtcaaa gaatgggtt tcaaccactg 113160
agacttcaaa gatagccac tgaaaaagtt ctaattcaca gaaattttcc aggaaaataa 113220
gaaataaaaaa gcacaaatga atatcttaa tttatccatg aaaataatg tgaaaatttta aactctaaga 113280
aaattgtaaa ataaagatata ttaatcattg agtacatata atatgcaaa agtgcattata 113340
ccatgtatata atgtgtgtac atatgtatgt gtacatata atatgcaaa agtgcattata 113400
gaaaccata attatataata atatgttagag ttccaagcat atctgcaggc tgaaggagag 113460
tctctgtgtt ttataggcat tgaagcattt ttagagtggta taggcaggta tataaatttt 113520
cacattgaaa ggaagaggaa gaaacagcta agtccctctc atgaggctt aagtggctt 113580
ggagagtcag ctttctctt cttcttcctt agtaaatttca atgcagacct tggcagccta 113640
atcactacaa tcctgcatacgt tgagtcctc tgaactgca ccatagttca ctagggaaaagg 113700
caactcatcca ccaagttcag agagggtggcc atccgcattt ctaaacagaa taaaacaaaa 113760
attcactttg tcatctgcct gacattggcc ttgctataag atagcctcat aaactgcctg 113820
aaaggcctta ttctttcaa aggattttgt gctgtctca agtgcatttt ttctccatgt 113880
cagaaataaa atatccagg attgacctgc ctcaaatttt agttgtctt gtttattggc 113940
cattgcattgt ataagtctaa ttgtgtgaca gaagttccg atactctgct tcatgaagta 114000
ccctttagga ctatctcctt gtctcttcag ttcttcataatc aatttttcctt gaagttctt 114060
tgtgaggctc tgatcacaaag agtaggcagg tgctggcatc cagcccagag ccacccttc 114120
cagccctagc acatccttaa agcaccctg accagcaccc tcaacctggc caacaggctt 114180
ttcagcatct acaggtaac ctccatggc ttgagctgca ctttcacaaa tcacaagaga 114240
aaccacaaaca cagagtacag ggaaaacagc tttcaaaaat tgtaagggtt agtagtggt 114300
ttaagatata accctgtctt cctattttagaa taccctattt ttttttgcc tgattgcct 114360
ggccagatct tccaatactg tgtttaatag gagtggtag agagggcatc cttgtctt 114420
gtcactttc aaaggaaatg ttccagctt ttgcccattt agtatgat tggctgtgag 114480
tttgcataaa atagctctt ttatttttagg atacatttca tcaataccta gtttatttgag 114540
agtttcagc atgaaggggt gttgaatttca taccatgtt tttttttctt catctattga 114600
gataatcatg tgatattttgt cattcgttct gtttatgtga tgattatgtt ttattgactt 114660
gggtatgtta agttagcctt gcatcccagc gataaagcca acttgatcgt ggtggataag 114720
cttttgcattt tgctgttggaa tttgggttgc cagtatttttta ttgaggatat ttgcatcgat 114780
gttcatcagg gatactggcc tgaaaattttc tttttttgtt gtgtctctgc cagggtttga 114840
tatcaggatg acgctggcct cataaaatga gtttagggagg agtcccttctt tttctgttat 114900
ttggaatagt ttccagaagga atggtaccag ctcctgtttt tacctctggt agaattcggc 114960
tgtgattcca tctggccctt gttttttttt gattggtagg ctattaatta ttgcacaaat 115020
ttcagaactt gttattgtatc tatccaggaa tttggcttct tccttgcattt ggagatgaca 115080
tgactgtata tttagaaaac cccattgtct cagccccaaa tctcccttaag ctgatgaaca 115140
tcttcagcaa agtcttagga tacaaaatca atgtgcaaaa atcacaagca ttcctataca 115200
ccaataacag acaaaacagag agacaaaatca tgagtgaattt cccattcaca attgctacaa 115260
agagaataaa atgccttagga atccaaacttca caagggatgt gaaggacctc ttcaaggaga 115320
actacaaacc accactcaag gaaataagag aggacacaaa cagatggaaa aacattccat 115380
gctcatggat aggaagaatc aatcatgtaa aaatggtcaa actgcccääa gtaatttatt 115440
gattcaatgc tatccctgtc aagctaccat tgactttctt cacagaatta gaaaatctc 115500
ttgtaaattt catatgaaac tgaaaaagaa cccatatacg caagacaatc ctaagaaaa 115560
agaactaagc tggaaagcattc atactacctg acttcaact atactacaag gctacagtaa 115620
caaaaacacg atgatactgg taccaaaaca catatataga ccaatggac agaacagagg 115680
cctcagaaat gatgccacac atctacaacc atctgatctt tgacaaaacct ggcaaaagca 115740
agcaatagga taaaaggattc cctgtttaat aaatgtatgtt gggaaaactg gctagccata 115800
tgcagaaaaac tgaagctgga ccccttcctt atgccttaga caaaaattaa ctcaagaggg 115860
ataaaaagact taaaacgtaa acctaaaagcc aaaaaacccc tagaagaaaa cctaggcaat 115920
accattcagg acataggaac aggccaaagac ttcatgacta aaacatcaa agcaatggca 115980
acagaagcca aaactggcaa atgggatctc actaaactaa agagcttctg cacagcaaag 116040
gaaactatca tcagagtggaa cagggacactt ccatccatcacttccatgaaatccat 116100
ccatccagaa tctataagga actttaacaa aaaaagtggg caaaagatata gaaacagacac 116160
aaaaatgtgg caaaagatata gaaacagacac ttccacagaag aagacatttca tggggcaat 116220
aaacatgtga aaaaacccatc atcatcaactg gtcatttagag aaatgcacat caaaaaaaaca 116280
caatggagata ccatctcactg ccagtttagaa tggcgatcat gaaaaagtca gggaaacaaca 116340
gatgctggag aggacgtggaa gaaacaggaa tgcttttaca ttgttagtgg gagtataat 116400
tagttcaacc attgtggaaag acagtgtgtt gattcctcgat ggtcttagaa ccggaaatac 116460
catttgcacc aacaatccca ttactgggtt acataccat aagattataa atattctact 116520
ataaaagacac atgcacatgt atgttcatttgc cagcactatt cacaatagca aagacttggaa 116580
accaacacaa atgcccattca accatagact ggataaaagaa aatatggcac atatacacca 116640
tggcacacta tgcagacata aaagaggatg agttcatgtc ctatgcaggc aaatggatga 116700
acctggaaac catcattctc agccaaacttca cacaggaaca gaaaacaaaaa cactgcattgt 116760

tgttcact tataagtggg tgtgaacaat gagaacacat ggacacagga agggaaat 116820
cacacaccag ggcttgtca gggttcgggg gctagggaaag ggatagcact aggagaaaata 116880
cttaatgtag attacaggt gatgggtgca gcaaaccacc atagcacata tataacctatg 116940
taacaaacct gcaggttctg cacatgtatc ccagaactta aagtgtata aaacaaatgt 117000
atgtcacctt gttaaaaaac agatagtaga aaaaccagca aatgagaac tggtgtctata 117060
taaaaatgtt tccaaagaac caagtaacat atgactttgt gctcaacttt aataatcatt 117120
agggaaatgtc acattacaat tacaagatat ttgtgtttt atatatctgt acctatatgt 117180
tgctaaatgtt caaaaagaaa aaaatgtat tgcataagtt tggtggact atgaaacaat 117240
tgaatctctc acataatgtt agtgagaata caaatggata caaccagttt gggaaaactat 117300
taggcagcat aaatttaagc tggacatatt tatactccat aagccagcac ttctactcat 117360
agttacagca gtcagacatg catgaatatg ttcatgaaga gatatgcaca agactgaatt 117420
ttgttaatgc cacatctgga aataacctat atataaagtt tggtgtactt attatacagc 117480
agtgcagatg aaggaattat ttacatattc atcacactca tcacatgaa taaaatcac 117540
caacataata ttttagcacca atgttagtgg cattggagac tagacctaac aaaatatgc 117600
ctgtataatt tctattacag aaaattcaaa accagacaca attaatctac agtgtaaaa 117660
gcaaagatat agttaatttt aggttactga ctagggaaagg gcatgaagac agatggtgga 117720
gtactgttaa tcttcttattt gagaattttag aagcaagtaa tacaagtatg ttcaactttgt 117780
aaaaattcat gaagcaggcc gggcacgggt gctcgtgcct gtaatcccag gactttggga 117840
ggccgaggca ggtgaatcac gaggtcagga ggttgagacc atcctggca gcacagtgaa 117900
accctgtctc tactaaaaat aaaaaaaatt agccaggcg ggtggcgggc acccgtagtc 117960
ccagctacta gggaggcgtga ggcaggagaa tggcatgaac ccggaggcg agcttcgagt 118020
gagccgagat cccaccactg cactccagcc tggcaacag agccacactc tgactaaaa 118080
aaaaaaaaaa aatcataatt catgaagctg tacacttacg ttttgtgctc attatttgc 118140
tgtaaatttag acttcaataat aaagcttact aaaaacgaat aaaaatagta ctgttctca 118200
agcaagcaaa gcttcatttc aatatcaaag cattcttattt acctatcagt acacagaggg 118260
tatttagttt ctagggtctg cacaataaag taccatgaac ttgtgtactt aaacatgcag 118320
atttatttcc tcacagttct agaggctaga agtccaagat caagggtgtgg gcaaaattgg 118380
tttcatctg agttcttctt ctggcttgc gatgatcatc ttatccgcac ctctttacac 118440
tttattttc tgggtgttattt tggattctaa tctcttctta taaggatgca agttatattg 118500
gattaggcga cagctcaccat attaggctttt attttactaa atggtctttagttagatttgc 118560
gtctccaaaca gtcatgctt gtgggtctg gagttagaac ttcagcatat gaattttgg 118620
gagggaggga aggggcacaa tccagttcat aacacagatt aagaaacgtg aaaggctaat 118680
agaagttga cacaatgtt gtgacactag tacgagagaa actgtatcag aaaagtggaa 118740
ttaagttgaa agtaacatgg taaacctaag gcaatgtgag aatccatggc agacatgaat 118800
gttatcttat ggattttcca atgtaaagaa gaaaataactg agaatgaaac ataaggcaga 118860
gaaggaccag agagttgtga gttccattt taaatttgc ttgtgccaat tgcataatct 118920
ctagagaaat tattcagtga gaaaaaaaaat ctgacagagt aattgtttca ttttgcata 118980
tctgtgaaat cccttagggaa aataaagtca tcatacaaat attataaattt attcctgtat 119040
ttgtcaccag aaaagccatt tgatattctt tggtaaggata gcttccct tattcataaa 119100
taagttctg catgtgtttt taatccctgga acacttgc tacaatcata tggatatttca 119160
gagttagata tatgattgtg atgattaaat gacttaggtg aaagaaaaat gccaatttacc 119220
agaaaaatgt agacacttag catttaaggt acttttattt gttaaagtct tgaataatga 119280
ggatggaaatg taatggcata aaaatataag aggcatgct taggatctt cactcaat 119340
aaatgaaagc taatatttt taagggtttt ccacacattt ggcacagtgc tatgcataatc 119400
acatacccca ttttgtgaaa tccggaaaaag agtgctttt tatttgcata tttcatctgt 119460
aatagaaaaa aattatagtc cagaattatt aagaaacatg acccagacta ctcagatcag 119520
aagttctgac atcagaatgt gaactcaacc agtgcactcc caaacgtatg tttctaccag 119580
tacagtatgc tttatgggtt gtagtggaaat ttcttctgt actaaccatg agggaaat 119640
gttattatcc atatcttattt taggcaaaat gtcatagaat tgggttggagg gttgaatgtg 119700
ttaaaaactta taaaatagat tagtgtttgg ctatataagaa acaccatgtt agtgcgtgtt 119760
aaatttagtgg taaaactaaa acacagaata aggaacatgt caaaaagaaca gagcagcatt 119820
tcagaatata ttaactccat atcctgtgaa ttgatattt gcttaaggctt tcatattttt 119880
atcaaggcat tcattttgtt gtttgtctt gtcctaaatc ttaagaatgtt caaccggatg 119940
tgtgcatttgc tcaaatacac aaagttggc acttttataa tatgaatttc actgtatata 120000
aattattgca taaaacaacat taaacattaa acaaagaaac aaaggtgaaa tctgacagga 120060
gcttgatata taaaatgaa tgaaggtatg tagggaaaag aaagagagat cagactgtt 120120
ctgtgtctgt gtagaaaggg aagacataag agactccatt ttgaaaaaaga cctgtacttt 120180
aaacaattgc tttgctgaga tttgtttaat ttgttagctt gccccagcca ctttgacc 120240
accacttgc cccaaacctgg agtcacaaa aacatgtt gatgaaatc aaggttaag 120300
tgatctaggg ctgtgcagga catgccttgc taacaaaatg tttacaagca gtatacattt 120360
gtaaaagtca tcgcatttc ttagtcttgc taaaccagga gcacaatgc ctgtggaaag 120420
ccgcaggggac ctctgcccctt gaaagcggag tattgtccaa ggtttctccc catgtgatag 120480
tctqaaatat ggtctqaaac caggggcaca ttgcactgcg gaaagccgtt gqqacctctg 120540

cccttggaaag cggggatttt tccaaaggttt ctccccatgt gatagtctga aatatggcct 120600
 tgtggcatga gaaagaccc accatcccc agcccgacac ccgttaaaggg tctgtatgt 120660
 ggaggatttt taaaagagga aagcttctt cagttgagat agaggaaggc cactgtctcc 120720
 tgcctgcccc tggggactga atgtctcggt ataaaacccg atagtagtattt ttttcaattc 120780
 tgagatcaga taaaaactgc cctatggtgg gaggtgagac acgtttgcag caatgctgcc 120840
 ttgttatttt ttaactccact gagatgtttt ggtggagaga aacataaaatc tggcttatgt 120900
 gcatgtccag tcatagttacc ttcccttggaa cttaaattatg acatagattt tattgtctcac 120960
 atgtttgttgc tgcacccctt tattatcacc ctgcctccctt actacattcc tttttgctga 121020
 aataatgaag ataataatc ataaaaactg agggaaactca gaggctgggt caggcccttg 121080
 atatgtctgg tggccggccc ctgggccccat tttttttttt ctgtactttt tctctgtgtc 121140
 ttattttttt tctcagtctt tcatccccacc caactagaaa taccacacagg tttttttttt 121200
 caggccaccc cttcaaaaggc acatcaactac tatgatttgc tagatgttgc cacagctttt 121260
 actcgatgtt gtataaaatca caaataaaaat ctttggttca gatattcaaa aataacccctt 121320
 tataagatgt tcttggtaat atagatgatt tttatataga aaaaatagaag gtacatttca 121380
 ttaatatcag tttttttttt tttttttttt gttgttaatgtt acagaacaca gctctgggta 121440
 tgctaagcaa aaccggaaaca ttctgaaaagg atgttagggc tccttaatca atgtgcgggt 121500
 aaaaagaccag gcttagagaa cagatagcag ctttcttcaatc gatattcaaa aataacccctt 121560
 aaccacacag aatcagtgtt tcttggtaac atcagcttca tcttcaaaacc tcactgttaat 121620
 gaatttagatg aaaaaaaaattt ttgtcaaaact tgagtcttta actcaactccc ttaagacaag 121680
 gagaggcttag catctgaccc ctgcttaagc ctttcttcaatc agagaatctc atctctcagt 121740
 gaaaaactaga agtggaaagat caagaggaca gaatttccat gactctcaac ttgacacattt 121800
 gaaaaattttc ctccaaaaaaa ttcaagtggaa gtgttaatgg ctggaaattttt atagctgtgc 121860
 aaccctcgaa acgcaaatat gttagatattt ttcttcaatc aaaaatggaaacaa tcatgcccctg 121920
 cctggaaaat ctcttatttgc ttgagtctgg aggaagaaga aagcaacaag tgagccccgc 121980
 ttgtttgctc aggcttctcc aagaggggat tgaactgggtt aaatcttatttca agaaataaca 122040
 agatcattca gtcaatattt tttttttttt tttttttttt tttttttttt tttttttttt 122100
 tctggccagaa ttataatcag gaataaaagaa atttttttttt tttttttttt tttttttttt 122160
 gcaagtccac ctataaaatgtt catttcttcaatc atataaaagag gggcagaata agagaaggag 122220
 aaagaagaga agaagaatag agacagaagc tgaatttagaa gggaaatgaaa agaaagacaa 122280
 agggacataa aacaagaaaa ataaaggcgtt tttttttttt tttttttttt tttttttttt 122340
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 122400
 gaaaagagggg gaaaactttt ggtaagaaga ttattactt actggtaata atagtaataa 122460
 tcatgttgcattt attgttatttgc aatcaatata tttttttttt tttttttttt tttttttttt 122520
 aattatgaga aggaacttca cgtttgggtttt tttttttttt tttttttttt tttttttttt 122580
 ccaggaattttt ctttcatgtt gttttttttt tttttttttt tttttttttt tttttttttt 122640
 agggaatctt agcatgtttttt cccttcatccca gttttttttt tttttttttt tttttttttt 122700
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 122760
 gatgttggggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 122820
 gtgttacccctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 122880
 tgacttagtac atggccatctt gttttttttt tttttttttt tttttttttt tttttttttt 122940
 gtgttccctt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123000
 catgttggggat tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123060
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123120
 agaccacca ctgatgtttttt tttttttttt tttttttttt tttttttttt tttttttttt 123180
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123240
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123300
 ctgttggctccatcttcatgtt tttttttttt tttttttttt tttttttttt tttttttttt 123360
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123420
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123480
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123540
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123600
 gttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123660
 cttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123720
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123780
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123840
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123900
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 123960
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 124020
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 124080
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 124140
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 124200
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 124260
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 124320

ccaatgcttt ggtccacagt gaggcacact tatgtttcca cagtccatat ctggtttatt 124380
 tcttggcttt ttagaggaaa aataaatgac cagtaggatt ctaaattgaa tgtaacctaa 124440
 aagtctgtct aaggagaccc aaggctaaca gataatctgt tcaatgcacg tctaaaacca 124500
 catcaaccca aaaggtaataa catgtgggtg taatgatctt gaagatgtgt gttttgggt 124560
 gtggaggcct atagttcccc tagaggtaac ctagaactta agatataatc acataacaagg 124620
 aagacgagag actaatgatg aggtgtgtgc aaggtagagaa attagaacag agatgcctc 124680
 aggaaattcg gattattgaa gggctgcaac ctctagagta ggacagagtc aacaacagca 124740
 acaacaagga tgaaactagt ccgtttaggc tttgttcat ggaggaggtt aaaagtctta 124800
 ttaagaatt tatgacaata gattaacggt cataattaatt cagcgtttat atttacactg 124860
 ctttgcgtgg ctatcaaattc tgaagctaaa aaaacatgaa ataggcccg cgtgggtgg 124920
 cacacgtgaa atcccagcac tttgggaggg ccaggcggtt ggatcacgag gtcaggagat 124980
 tgagaccatc tgcctaacac ggtgaaaccc cgtctctact aagagtacaa aaaatcagca 125040
 gggcgtgggt gccccgcggc gtagtcttag ctactcagga gactgaggca gaagaatggc 125100
 ctgaacttgg gaggcgcagc ttgcagttag tggagatgtt gcoactgcac tccagcctgg 125160
 gtgacagaag gagactctgt ctcaaaacaa acataaaacaa aaacaaaaac aagcaaaacaa 125220
 aaaacaaaaaa caagcaaaca aaaaacaaaa catabaggg cccaggttag aagaaacctc 125280
 atgacactgg aataagcaaa tataaacata agtcagagaa tgaattatca actactcaga 125340
 cttcaaaagat atcccgcaga taagttctt attgacagac attttccagg aaaataagaa 125400
 ataaaaagca caaatgaata cttcaactt attacagata ttaggattac ttgatgcaaa 125460
 ttataaaaata aagatattt atcattgaaa taagtgtgaa aattaaaatt cgaagaccat 125520
 atatatgtgt gtatatatat gtatgtgtac atatataatgc aaagactgtat tatataaagc 125580
 aataattata tataatatgt agagttccga gcatatctgc agggcgaagg agagactctg 125640
 tttttatag gcattgaagc attttacatt ttttttatg ttctcagcac ctttatttt 125700
 ttattttttt tttagtatt tattgatcat tcttgggtgt ttctcgaaga gggggattt 125760
 gcagggtcat aggacaatag tggagggaaag gtcagcagat aaacatgtga acaagggtct 125820
 ctggtttcc taggcagagg accctgcggc cttccgcgtt gttgtgtcc ctgggtactt 125880
 gagattaggg agtgggtatg actcttaacg agcatgcgc cttcaagcat ctgtttaaca 125940
 aagcacatct tgcaccgcggc ttaatccatt taaccctgag tggacacagc acatgtttca 126000
 gagagcacgg ggttgggggt aagttatag attaacagca tcccaaggca gaagaatttt 126060
 tcttagtaca gaacaaaatg gagtctctt tgcactactt tttctacaca gacacagtaa 126120
 caatctgatt tcttttttcc ttccccacat ttccccctta tctatggac aaaactgcac 126180
 tcgtcatcat ggcccattt caatgagctg ttgggtacac ctcccagacg ggggtggcggc 126240
 cggcgaggg ggctcctcac ttcccagacg gggggggcgg gcaagaggcgc cccccacctc 126300
 cgggggggggg tggctgtgg gcggggggctg cccccaccc ccctccggc tggggcggct 126360
 ggcgggttag gggctgcccc ccacccctt cccggacggg gcaagctggc gggcgggggc 126420
 tgcgcgcgc ctcctcccg gacggggcag ctggccggc aggggctgac ccccacctcc 126480
 tggacagggc ggctgtggg tggagacacg ctcacttcc cggatggggc ggctgccagg 126540
 cggaggggct ctcacttcc cagacgtggc ggctgccccg cggaggggct ctcacttct 126600
 cagatgtggc ggccggccatg cggaggagct ctcacttcc caggcagggc agccgggcag 126660
 agacggtcct cacctcccg acgggggtggc ggtcggggcag agacacttcc cagttcccg 126720
 acgggggtgc gccaaggcag aggcgttcc cacttcccg actgggtggc cgagcagagg 126780
 ggctccctac atcccagacg atggggggcc aggcagagac gtccttcaact tcccagacag 126840
 ggtggggcgt gggcagaggc tgcaatctgg gcactttggg aggccaaggc aggagctgg 126900
 gaggtggagg ttgttagcaag ccgagatgac gccactgcac tccagctgg gtaacattga 126960
 gcactgagtg agtggagact cgtctgcaat cccggcaccc cggggggctg aggccggcag 127020
 atcaactcgcg gtcaggagct ggagaccagc cggccaaaca cggcgaaacc cacaaaaaaa 127080
 tacaaaaacc agtcagggtga gcggggcatg ctcgaatcc cggcactcg gcaggctgag 127140
 gcaggagaat caggcaggga ggttgcgtt agcccgatg gcccgcgtac agtccagcct 127200
 cggctgggca tcagagggat accgtggaga gagagggaga gggagaggaa gagggagacc 127260
 gtggaggggag agggagaggaa agaccgtggc aggagaggaa gagggagaggc attgaagcat 127320
 ttttagatgg ataggttagt atataaattt tcacattgaa aggaagaggaa agaaacagct 127380
 aagtccctct catgagactg tcttggagag ggttacagca gagaggctgc ctttctctcc 127440
 ttcctccctca gtaaaatttc tacagacctt ggcagcccaa tctctacagt cctgcagcag 127500
 tgacttgcctc gaactggcac catagttcac tgggaaaggc attcatccat caagttcaga 127560
 gaggtggcca tccatgttgc taaacagaat aaaacaaaaa ttcactttgt catctgcctg 127620
 acattggcct tgctatgcaa tagcatcata aactgcctga aaggctttat tattttcaaa 127680
 ggattttgtg ctgtctacaa gtggtatttt tctccaccc agaaataaaa tatttttagga 127740
 ttgaccacc tcaaatttgc gtttgtctt gttatcgcc actgcttata taagtctaat 127800
 tgtgtgacag aagttcccaa tgctctgtt catgaagtg cctttaggac tacctcttgc 127860
 ctcttcagtt cttcaatcaa cttttcctg aagttctgt gtggggatct gatataaaca 127920
 gtaggcagct gctggcatcc agcccagagc cacctctcc agggccagga catccttaaa 127980
 gcacctctga ccagcaacct caatctggcc aacactttt tcagcatata caggtaaatc 128040
 tccatgggct tgatcagcac cttcacaaat cagaaaagaa accacaaccc agagtaaaga 128100

aaaaacaacc tttcaaagg ttggtgggtg taaaatatca cccggtaaa 128160
aaacaaacaa actgatagca gaaaaaaatg gcaaaatatg aactggtgct atataaaaac 128220
atatttaag aaccaagtaa cttatgactt tgcgtcaac ctgaataata attagggaaag 128280
tgtacattac aagatatctg tatctttata tatctgtacc tatataatctg tgcgtatg 128340
ttgcagaatg acagaaaaac tgatattgtt aagcattgga gagactgtgg aacaactgac 128400
cctctcatat aatgcttagt ggaatgcaaa tggatataaa cagtatggaa aactactagg 128460
cagcataaaat ttaagctgca catatttata ctccataaaac caacactact gcccggtaaa 128520
ctaacaggca gaaatacatg aatatgttcg tgaagagata tgacacaagat tgaattttgt 128580
aacaggcaca cctggaaagta tcctatatgt gtatcaatag taggaaggat ttattaat 128640
tggtgactt actatacagc cgtggagatg aaggaaattat ttacatattc atgacactcc 128700
tcattttgaa taaaatcac caacataata tttagcacca atgtatgtt cattggaaagg 128760
ctagaccta cagaatatgc actgtataat ttcttattatg gaaaactcaa aagcagacac 128820
aattaatgt a cagtgttaaa agtgaagatg tagttaattt gaggttagt actaggaaag 128880
ggcaaggaga gagatagtgc agtactgtt atgttctact tgaggattt gaggcaagta 128940
atacaagtat gttcaacttg taaaattca tgaagctgta cgccctacatt ttgtgcacac 129000
tttctgtgtg taaatttagac ttcaatataa aggttactaa aaacgaataa aaatagtact 129060
agacttcaag caagtaaagc ttcatccaa tatcaagca ttcttattac ccatcgtac 129120
acagagggtt ttagttgtc agggctgcca caaataagta ccatgaactt ggtgacttaa 129180
acatgcagat ttatccctc acagttctag aaaactgggtt tcattctgag ttcttttct ctagcataatc tcttcttata aggtgcaag 129240
ctttacactt tcttttctg tgcgtatctg tcatatttgc tttagggcaca gctcacccgc tggggcttgg agctgaaact tcagcataag 129300
cgatatcgtg tctccaaacag tcacgctctg aattttggag agtgaggaaa gaggcacaat aatgctaata gaattttgac acaaagttt 129360
aaagttgaat taagttgaaa gtaacatggt tgacacttgc tagagaaaattt atccatggca 129420
gtcaggaatg ttattgtatg gattttctaa taaggcagag aaggaccaga gatgtgttagt ttttgcataatc ctgtgaaatc ttccatgtt 129480
ttcctgtatt tgcaccaga aaagccattt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 129540
ttcataggtt agtttctgca tgcgtatctt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 129580
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 129640
ttcctgtatt tgcaccaga aaagccattt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 129700
ttcataggtt agtttctgca tgcgtatctt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 129760
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 129840
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 129900
ttcctgtatt tgcaccaga aaagccattt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 129960
ttcataggtt agtttctgca tgcgtatctt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 130020
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130080
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130140
ttcctgtatt tgcaccaga aaagccattt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 130200
ttcataggtt agtttctgca tgcgtatctt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 130260
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130320
ttcctgtatt tgcaccaga aaagccattt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 130380
ttcataggtt agtttctgca tgcgtatctt atgtatccatc agtttctgca tgcgtatctt aaaaatatac taacagagta attgcttcat 130440
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130500
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130560
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130620
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130680
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130740
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130800
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130860
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130920
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 130980
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131040
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131100
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131160
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131220
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131280
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131340
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131400
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131460
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131520
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131580
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131640
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131700
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131760
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131820
tttgcataatc ctgtgaaatc ctttagggaa tgcgtatctt aaaaatatac taacagagta attgcttcat 131880

gagcctggag gaataagaaa gcaacaaatg agcccaacct gtttgcctag gcttcagg 131940
 gaaggatcg gattggtcaa aactattcac aaagaacaag atcctcgccc aaatattgct 132000
 ttttatact tgccttttc tctctttacg cgcattctct gcccagtatt ataatcgaaa 132060
 aaaaagaaat tgcttagttc tgaatgttat gtttggggc aagtccacct ataaagtaca 132120
 tttctcaaat aaaaggagag gcagaataag gggaggagaa acgagagaag aagaatagaa 132180
 aaacagaagc tgaatttcaa gggagtgaaa agaaagaaaa tggaaagaaaa acaagaaaa 132240
 taaaggcgag tatacataaaa ggttagagac gaaacaaagt tcagtagaaa tagaatgagg 132300
 taaaagatca ttactgtat agtaataatc atgttaagtt ctttcttcc accgaaggaa
 gagttcattc tcctggact gaccagtcgc 132540
 tttctggcca tttacatggt cacgggtggca 132600
 gccaacgcct ggctccacat gcccattgtac 132660
 ctgtgtttct cttccaaatgt gactccaaag 132720
 agcatttcct atcctgcctg tcttgcag 132780
 gagatctaca tcctggctgt gatggcctt 132840
 ctttatggca gcagaatgtc caagagtgt 132900
 tatggagcgc tcaactggcct gatggagacc 132960
 cccaatgaaa ttaatcactt ctactgtgcg 133020
 gacacccata acaaggagtt gtcaatgtt 133080
 ctcttcatca tatgtatttc ctacccattac 133140
 acagagggca ggcaaaaaagc ttttcttacc 133200
 ttctatgcaa ccctttctt catgtatctc 133260
 ggtaaaatgg tagctgtatt ttataccaca 133320
 agccttagaa ataaaaatgt aaaagaagca 133380
 ttttcttaaa aatcagttt ctttgggtt 133440
 ctgagaaata tagtgcatca atggagaaca 133500
 ttttattatt atattttagt atggagttc 133560
 gatctcggct cactgcaacc tttgcctccc 133620
 ccgagtagct gggAACACAA ggcacaccca 133680
 agagacaggg tttcaccatg ctggccaggc 133740
 atggaaacacc ggcggccagtc taaaacttt 133800
 gttcacaaaa gctttatgt ttaagttgt 133860
 tctttgtctt cctgtgcacag aatggcttt 133920
 catgtggcta tttctggaaatgagattaa 133980
 ctttaattgtatgttgc ttttgagtg 134040
 ccagcaagac ttaatataatg ggctgcccag 134100
 actcacaatg gcaatacataa acggctgata 134160
 tggagttgc ttgtttccctt agcacaatataa 134220
 tacaatttca cacaggctga attatataat 134280
 aatcacaattt gagattttgt tttgtgttct 134340
 atgtgtgttc agttaaaattt atttattata 134400
 tatgtatgtt taaatgtcaa ataattttac 134460
 accttaccac taactccccca tagataatac 134520
 attttaattt atatgtttt attcctctgt 134580
 tgacatctgt ttattaaaga taagggtcc 134640
 cctacaatga tcacttctt ctctcatttgc 134700
 ttaatcagta tctactgttta aaattactat 134760
 ctataactat ttgagcatca ctaacactat 134820
 tgattgtat tcttttcttgc cacagttt 134880
 gttttgtctt tttttttgtt tagtagctac 134940
 caccttataa gggtaaaaca caatagatga 135000
 ttctctgtcc tcccatttggat ttctgtctg 135060
 tccaatcatt ggtttaaaggc atttattca 135120
 aactggagat tcaacaaaaa taagacagaa 135180
 ttaatgaact tttctgtgtt tattatgttgc 135240
 catgtttgtt attttctgttgc atgtcagaat 135300
 atgagactaca tggggttccc gagggggca 135360
 gtcagagcca tgggcaaaaca ggctcaagtc 135420
 tgaattacag tatatttaat cagttatataa 135480
 aaagtaacat ttatcatcaa gaataaagg 135540
 gagagtgaca tgaacaaaaa gaatagccta 135600
 gtaaagaaga gttttcgatg tggcaaaagc 135660

gcttcatagt ataactgttc atcttcagtt acagaactgc tactgagata acataactaa 143280
 agcctttgg ctcttttat acaaaggatg atatthaact agggttttag tgattttaa 143340
 aaagtttctc tttcttcctt gatattcaga ccaatgcgtc tcatatgaga tgaagaaatg 143400
 tccagaagaa actatactga actgacagaa tttgttctct tgggtctaac aagccgtcca 143460
 gagctgcgag ttgctttctt ggcactgttc cttttgtct acatagccac tgtggtagga 143520
 aacttgggaa tgattattt aatcaaagtt gattctcgac ttacactcc catgttaattt 143580
 tttcttcctt gtttgcctt tctagatctg ttttcttcca caaatttcac tcccaaaatg 143640
 ctagaaaatt tcttatcaga gaagaagacc atttcctatg caggttgttt gatgcagtgc 143700
 tatgttgtca ttgctgtgtt cttgcagag cactgcatgt tggcagtcat ggcataatgac 143760
 cgctatatgg ccatctgtaa tccattgctc tacagtagca aatgtccca aggtgtttgt 143820
 gtccacactgg tcattgtccc ttatgtctat ggcttcttc tcagtgtat gaaacctta 143880
 aggacccata acctctcctt ctgtggaca aatgaaatca accatttcta ctgtgtat 143940
 cctccttta tcaaactggc atgcctgc acgtacagca aggagctgtc catgtacata 144000
 gtagccggct acagcaacgt ccagtctt cttatcattc tccatccata catgttcata 144060
 cttgtcgcta tcctcagaag ccattctgc gaggaaagga aaaaagctt ttccacatgt 144120
 ggttcccacc tgacagttgtt cacaatctt tatggAACCC tttctgtat gcatttgaga 144180
 cgtcccacag acgagtcgtt ggagcagggg aatgtgtgg ctgtgtttt caccacagt 144240
 atactcatgc tgaactccat gatctatggc ctcaggaaca agatgtgaa agaggcggt 144300
 aaaaaagcaa tagggaaaaca aacattgggaa aataaaaaat gctaagctat cattaaaaat 144360
 ttgtgaagta atgagatata atatcattgg gttagatgtc acattttagg ctacatttgc 144420
 acaattcatt tctaatttgc tgtaggtat ctgactgagt 144460

<210> 2
 <211> 195
 <212> DNA
 <213> Homo sapiens

<400> 2
 atg agc ttc tta ata aga agt gat tca aca cta cac act cca atg tgc 48
 Met Ser Phe Leu Ile Arg Ser Asp Ser Thr Leu His Thr Pro Met Cys
 1 5 10 15
 ttg ttc ctc agt cat ctc tcc ttt gta gat ctc tat tat gcc acc aat 96
 Leu Phe Leu Ser His Leu Ser Phe Val Asp Leu Tyr Tyr Ala Thr Asn
 20 25 30
 gcc act cct ccg atg ctg gtt aac ttt ttt cca aga gaa aaa ccg 144
 Ala Thr Pro Pro Met Leu Val Asn Phe Phe Phe Pro Arg Glu Lys Pro
 35 40 45
 ttt cct tta ttg gtt gct tta tcc aat ttc acc ttt tca ttg cac tgg 192
 Phe Pro Leu Leu Val Ala Leu Ser Asn Phe Thr Phe Ser Leu His Trp
 50 55 60
 tga 195
 *
 65

<210> 3
 <211> 948
 <212> DNA
 <213> Homo sapiens

<400> 3
 atg ttc tcc cca aac cac acc ata gtg aca gaa ttc att ctc ttg gga 48
 Met Phe Ser Pro Asn His Thr Ile Val Thr Glu Phe Ile Leu Leu Gly
 1 5 10 15
 ctg aca gac gac cca gtg cta gag aag atc ctg ttt ggg gta ttc ctt 96
 Leu Thr Asp Asp Pro Val Leu Glu Lys Ile Leu Phe Gly Val Phe Leu
 20 25 30
 gcg atc tac cta atc aca ctg gca ggc aac ctg tgc atg atc ctg ctg 144
 Ala Ile Tyr Leu Ile Thr Leu Ala Gly Asn Leu Cys Met Ile Leu Leu
 35 40 45
 atc agg acc aat tcc cac ctg caa aca ccc atg tat ttc ttc ctt ggc 192
 Ile Arg Thr Asn Ser His Leu Gln Thr Pro Met Tyr Phe Phe Leu Gly
 50 55 60
 cac ctc tcc ttt gta gac att tgc tat tct tcc aat gtt act cca aat 240

His	Leu	Ser	Phe	Val	Asp	Ile	Cys	Tyr	Ser	Ser	Asn	Val	Thr	Pro	Asn	
65															80	
atg	ctg	cac	aat	ttc	ctc	tca	gaa	cag	aag	acc	atc	tcc	tac	gct	gga	288
Met	Leu	His	Asn	Phe	Leu	Ser	Glu	Gln	Lys	Thr	Ile	Ser	Tyr	Ala	Gly	
															95	
tgc	ttc	aca	cag	tgt	ctt	ctc	ttc	atc	gcc	cta	gtg	atc	act	gag	ttt	336
Cys	Phe	Thr	Gln	Cys	Leu	Leu	Phe	Ile	Ala	Leu	Val	Ile	Thr	Glu	Phe	
															110	
tac	ttc	ctt	gct	tca	atg	gca	ttg	gat	cgc	tat	gta	gcc	att	tgc	agc	384
Tyr	Phe	Leu	Ala	Ser	Met	Ala	Leu	Asp	Arg	Tyr	Val	Ala	Ile	Cys	Ser	
															115	
cct	tta	cat	tac	agt	tcc	agg	atg	tcc	aag	aac	att	tgc	atc	tct	ctg	432
Pro	Leu	His	Tyr	Ser	Ser	Arg	Met	Ser	Lys	Asn	Ile	Cys	Ile	Ser	Leu	
															130	
gtc	act	gtg	cct	tac	atg	tat	ggc	ttc	ctt	aat	ggg	ctc	tct	cag	aca	480
Val	Thr	Val	Pro	Tyr	Met	Tyr	Gly	Phe	Leu	Asn	Gly	Leu	Ser	Gln	Thr	
															145	
ctg	ctg	acc	ttt	cac	tta	tcc	ttc	tgt	ggc	tcc	ctt	gaa	atc	aat	cat	528
Leu	Leu	Thr	Phe	His	Leu	Ser	Phe	Cys	Gly	Ser	Leu	Glu	Ile	Asn	His	
															165	
ttc	tac	tgc	gct	gat	cct	cct	atc	atg	ctg	gcc	tgc	tct	gac	acc		576
Phe	Tyr	Cys	Ala	Asp	Pro	Pro	Leu	Ile	Met	Leu	Ala	Cys	Ser	Asp	Thr	
															180	
cgt	gtc	aaa	aag	atg	gca	atg	ttt	gta	gtt	gca	ggc	ttt	act	ctc	tca	624
Arg	Val	Lys	Lys	Met	Ala	Met	Phe	Val	Val	Ala	Gly	Phe	Thr	Leu	Ser	
															195	
agc	tct	ctc	ttc	atc	att	ctt	ctg	tcc	tat	ctt	ttc	att	ttt	gca	gcg	672
Ser	Ser	Leu	Phe	Ile	Ile	Leu	Leu	Ser	Tyr	Leu	Phe	Ile	Phe	Ala	Ala	
															210	
atc	ttc	agg	atc	cgt	tct	gct	gaa	ggc	agg	cac	aaa	gcc	ttt	tct	acg	720
Ile	Phe	Arg	Ile	Arg	Ser	Ala	Glu	Gly	Arg	His	Lys	Ala	Phe	Ser	Thr	
															225	
tgt	gct	tcc	cac	ctg	aca	ata	gtc	act	ttg	ttt	tat	gga	acc	ctc	ttc	768
Cys	Ala	Ser	His	Leu	Thr	Ile	Val	Thr	Leu	Phe	Tyr	Gly	Thr	Leu	Phe	
															245	
tgc	atg	tac	gta	agg	cct	cca	tca	gag	aag	tct	gta	gag	gag	tcc	aaa	816
Cys	Met	Tyr	Val	Arg	Pro	Pro	Ser	Glu	Lys	Ser	Val	Glu	Glu	Ser	Lys	
															260	
ata	act	gca	gtc	ttt	tat	act	ttt	ttg	acc	cca	atg	ctg	aac	cca	ttg	864
Ile	Thr	Ala	Val	Phe	Tyr	Thr	Phe	Leu	Thr	Pro	Met	Leu	Asn	Pro	Leu	
															275	
atc	tat	agc	cta	cg	aa	aca	gat	gta	atc	ctt	gcc	atg	caa	caa	atg	912
Ile	Tyr	Ser	Leu	Arg	Asn	Thr	Asp	Val	Ile	Leu	Ala	Met	Gln	Gln	Met	
															290	
att	agg	gga	aaa	tcc	ttt	cat	aaa	att	gca	gtt	tag					948
Ile	Arg	Gly	Lys	Ser	Phe	His	Lys	Ile	Ala	Val	*					
															305	
															310	
															315	

<210> 4
<211> 519
<212> DNA
<213> Homo sapiens

<400> 4																		
atg	tta	aag	aaa	aac	cat	aca	gcc	gtg	act	gag	ttt	gtt	ctc	ctg	gga		48	
Met	Leu	Lys	Lys	Asn	His	Thr	Ala	Val	Thr	Glu	Phe	Val	Leu	Leu	Gly			
1															5	10	15	
ctg	aca	gat	cg	g	ct	g	at	cc	tt	tt	tt	tt	gt	tt	ct		96	
Leu	Thr	Asp	Arg	Ala	Glu	Leu	Gln	Ser	Leu	Leu	Phe	Val	Val	Phe	Leu			
															20	25	30	
gtc	atc	tac	ctt	atc	aca	gta	atc	ggc	aat	gtg	agc	atg	atc	ttg	tta		144	
Val	Ile	Tyr	Leu	Ile	Thr	Val	Ile	Gly	Asn	Val	Ser	Met	Ile	Leu	Leu			
															35	40	45	

atc aga agt gac tcg aca cta cac act cca atg tac ttc ttc ctc agt	192
Ile Arg Ser Asp Ser Thr Leu His Thr Pro Met Tyr Phe Phe Leu Ser	
50 55 60	
cac ctc tcc ttt gta gat ctc tgt tat acc acc aat gtt act cct cag	240
His Leu Ser Phe Val Asp Leu Cys Tyr Thr Asn Val Thr Pro Gln	
65 70 75 80	
atg ctg gtt aac ttt tta tcc aag aga aaa acc att tcc ttc atc ggc	288
Met Leu Val Asn Phe Leu Ser Lys Arg Lys Thr Ile Ser Phe Ile Gly	
85 90 95	
tgc ttt atc caa ttt cac ttt ttc att gca ctg gtg att aca gat tat	336
Cys Phe Ile Gln Phe His Phe Ile Ala Leu Val Ile Thr Asp Tyr	
100 105 110	
tat atg ctc aca gtg atg gct tat gac cgc tac atg gcc atc tgc aag	384
Tyr Met Leu Thr Val Met Ala Tyr Asp Arg Tyr Met Ala Ile Cys Lys	
115 120 125	
ccc ttg tta tat gga agc aaa atg acc agg tgt gtc tgc ctc tgt ctg	432
Pro Leu Leu Tyr Gly Ser Lys Met Thr Arg Cys Val Cys Leu Cys Leu	
130 135 140	
gct gct gct ccc tat att tat ggc ttt gca aat ggt cta agc aca gac	480
Ala Ala Ala Pro Tyr Ile Tyr Gly Phe Ala Asn Gly Leu Ser Thr Asp	
145 150 155 160	
cac cct gat gct tcg tct gtc ctt ctg tgg acc caa tga	519
His Pro Asp Ala Ser Ser Val Leu Leu Trp Thr Gln *	
165 170	

<210> 5

<211> 948

<212> DNA

<213> Homo sapiens

<400> 5	
atg ttg tcc cca aac cac acc ata gtg aca gaa ttc att ctc tta gga	48
Met Leu Ser Pro Asn His Thr Ile Val Thr Glu Phe Ile Leu Leu Gly	
1 5 10 15	
ctg aca gac gac cca gtg cta gag aag atc ctg ttt ggg gtg ttc ctg	96
Leu Thr Asp Asp Pro Val Leu Glu Lys Ile Leu Phe Gly Val Phe Leu	
20 25 30	
gcg atc tac cta atc aca ctg gca ggc aac ctg tgc atg atc ctg ctg	144
Ala Ile Tyr Leu Ile Thr Leu Ala Gly Asn Leu Cys Met Ile Leu Leu	
35 40 45	
atc agg acc aat tcc caa ctg caa aca ccc atg tat ttc ttc ctt ggt	192
Ile Arg Thr Asn Ser Gln Leu Gln Thr Pro Met Tyr Phe Leu Gly	
50 55 60	
cac ctc tcc ttt tta gac att tgc tat tct tcc aat gtt act cca aat	240
His Leu Ser Phe Leu Asp Ile Cys Tyr Ser Ser Asn Val Thr Pro Asn	
65 70 75 80	
atg ctg cac aat ttc ctc tca gaa cag aag acc atc tcc tac gct gga	288
Met Leu His Asn Phe Leu Ser Glu Gln Lys Thr Ile Ser Tyr Ala Gly	
85 90 95	
tgc ttc aca cag tgt ctt ctc atc gcc cta gtg atc act gag ttt	336
Cys Phe Thr Gln Cys Leu Leu Phe Ile Ala Leu Val Ile Thr Glu Phe	
100 105 110	
tac ttc ctt gct tca atg gca ttg gat cgc tat gta gcc att tgc agc	384
Tyr Phe Leu Ala Ser Met Ala Leu Asp Arg Tyr Val Ala Ile Cys Ser	
115 120 125	
cct tta cat tac agt tcc agg atg tcc aag aac att tgc atc tct ctg	432
Pro Leu His Tyr Ser Ser Arg Met Ser Lys Asn Ile Cys Ile Ser Leu	
130 135 140	
gtc act gtg cct tac atg tat ggc ttc ctt aat ggg ctc tct cag aca	480
Val Thr Val Pro Tyr Met Tyr Gly Phe Leu Asn Gly Leu Ser Gln Thr	
145 150 155 160	
ctg ctg acc ttt cac tta tcc ttc tgt ggc tcc ctt gaa atc aat cat	528
Leu Leu Thr Phe His Leu Ser Phe Cys Gly Ser Leu Glu Ile Asn His	

49

165	170	175	
ttc tac tgc gct gat cct cct ctt atc atg ctg gcc tgc tct gac acc			576
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Met Leu Ala Cys Ser Asp Thr			
180	185	190	
cgt gtc aaa aag atg gca atg ttt gta gtt gca ggc ttt act ctc tca			624
Arg Val Lys Lys Met Ala Met Phe Val Val Ala Gly Phe Thr Leu Ser			
195	200	205	
agc tct ctc ttc atc att ctt ctg tcc tat ctt ttc att ttt gca gcg			672
Ser Ser Leu Phe Ile Ile Leu Leu Ser Tyr Leu Phe Ile Phe Ala Ala			
210	215	220	
atc ttc agg atc cgt tct gct gaa ggc agg cac aaa gcc ttt tct acg			720
Ile Phe Arg Ile Arg Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr			
225	230	235	240
tgt gct tcc cac ctg aca ata gtc act ttg ttt tat gga acc ctc ttc			768
Cys Ala Ser His Leu Thr Ile Val Thr Leu Phe Tyr Gly Thr Leu Phe			
245	250	255	
tgc atg tac gta agg cct cca tca gag aag tct gta gag gag tcc aaa			816
Cys Met Tyr Val Arg Pro Pro Ser Glu Lys Ser Val Glu Glu Ser Lys			
260	265	270	
ata att gca gtc ttt tat act ttt ttg agc cca atg ctg aac cca ttg			864
Ile Ile Ala Val Phe Tyr Thr Phe Leu Ser Pro Met Leu Asn Pro Leu			
275	280	285	
atc tat agc cta cgg aac aga gat gta atc ctt gcc ata caa caa atg			912
Ile Tyr Ser Leu Arg Asn Arg Asp Val Ile Leu Ala Ile Gln Gln Met			
290	295	300	
att agg gga aaa tcc ttt tgt aaa att gca gtt tag			948
Ile Arg Gly Lys Ser Phe Cys Lys Ile Ala Val *			
305	310	315	

<210> 6

<211> 918

<212> DNA

<213> Homo sapiens

<400> 6

atg tcc aac aca aat ggc agt gca atc aca gaa ttc att tta ctt ggg		48	
Met Ser Asn Thr Asn Gly Ser Ala Ile Thr Glu Phe Ile Leu Leu Gly			
1	5	10	15
ctc aca gat tgc ccc gaa ctc cag tct ctg ctt ttt gtg ctg ttt ctg		96	
Leu Thr Asp Cys Pro Glu Leu Gln Ser Leu Leu Phe Val Leu Phe Leu			
20	25	30	
gtt gtt tac ctc gtc acc ctg cta ggc aac ctg ggc atg ata atg tta		144	
Val Val Tyr Leu Val Thr Leu Leu Gly Asn Leu Gly Met Ile Met Leu			
35	40	45	
atg aga ctg gac tct cgc ctt cac acg ccc atg tac ttc ttc ctc act		192	
Met Arg Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Thr			
50	55	60	
aac tta gcc ttt gtg gat ttg tgc tat aca tca aat gca acc ccc cag		240	
Asn Leu Ala Phe Val Asp Leu Cys Tyr Thr Ser Asn Ala Thr Pro Gln			
65	70	75	80
atg tcg act aat atc gta tct gag aag acc att tcc ttt gct ggt tgc		288	
Met Ser Thr Asn Ile Val Ser Glu Lys Thr Ile Ser Phe Ala Gly Cys			
85	90	95	
ttt aca cag tgc tac att ttc att gcc ctt cta ctc act gag ttt tac		336	
Phe Thr Gln Cys Tyr Ile Phe Ile Ala Leu Leu Thr Glu Phe Tyr			
100	105	110	
atg ctg gca gca atg gcc tat gac cgc tat gtg gcc ata tat gac cct		384	
Met Leu Ala Ala Met Ala Tyr Asp Arg Tyr Val Ala Ile Tyr Asp Pro			
115	120	125	
ctg cgc tac agt gtg aaa acg tcc agg aga gtt tgc atc tgc ttg gcc		432	
Leu Arg Tyr Ser Val Lys Thr Ser Arg Arg Val Cys Ile Cys Leu Ala			
130	135	140	
aca ttt ccc tat gtc tat ggc ttc tca gat gga ctc ttc cag gcc atc		480	

50

Thr	Phe	Pro	Tyr	Val	Tyr	Gly	Phe	Ser	Asp	Gly	Leu	Phe	Gln	Ala	Ile	
145					150					155					160	
ctg	acc	ttc	cgc	ctg	acc	ttc	tgt	aga	tcc	aat	gtc	atc	aac	cac	ttc	528
Leu	Thr	Phe	Arg	Leu	Thr	Phe	Cys	Arg	Ser	Asn	Val	Ile	Asn	His	Phe	
					165				170				175			
tac	tgt	gct	gac	ccg	ccg	ctc	att	aag	ctt	tct	tgt	tct	gat	act	tat	576
Tyr	Cys	Ala	Asp	Pro	Pro	Leu	Ile	Lys	Leu	Ser	Cys	Ser	Asp	Thr	Tyr	
					180				185				190			
gtc	aaa	gag	cat	gcc	atg	ttc	ata	tct	gct	ggc	ttc	aat	ctc	tcc	agc	624
Val	Lys	Glu	His	Ala	Met	Phe	Ile	Ser	Ala	Gly	Phe	Asn	Leu	Ser	Ser	
					195				200				205			
tcc	ctc	acc	atc	gtc	ttg	gtg	tcc	tat	gcc	ttc	att	ctt	gct	gcc	atc	672
Ser	Leu	Thr	Ile	Val	Leu	Val	Ser	Tyr	Ala	Phe	Ile	Leu	Ala	Ala	Ile	
					210				215				220			
ctc	cg	atc	aaa	tca	gca	gag	gga	agg	cac	aag	gca	ttc	tcc	acc	tgt	720
Leu	Arg	Ile	Lys	Ser	Ala	Glu	Gly	Arg	His	Lys	Ala	Phe	Ser	Thr	Cys	
					225				230				235		240	
ggt	tcc	cat	atg	atg	gct	gtc	acc	ctg	ttt	tat	ggg	act	ctc	ttt	tgc	768
Gly	Ser	His	Met	Met	Ala	Val	Thr	Leu	Phe	Tyr	Gly	Thr	Leu	Phe	Cys	
					245				250				255			
atg	tat	ata	aga	cca	cca	aca	gat	aag	act	gtt	gag	gaa	tct	aaa	ata	816
Met	Tyr	Ile	Arg	Pro	Pro	Thr	Asp	Lys	Thr	Val	Glu	Glu	Ser	Lys	Ile	
					260				265				270			
ata	gct	gtc	ttt	tac	acc	ttt	gtg	agt	ccg	gta	ctt	aat	cca	ttg	atc	864
Ile	Ala	Val	Phe	Tyr	Thr	Phe	Val	Ser	Pro	Val	Leu	Asn	Pro	Leu	Ile	
					275				280				285			
tac	agt	ctg	agg	aat	aaa	gat	gtg	aag	cag	gcc	ttg	aag	aat	gtc	ctg	912
Tyr	Ser	Leu	Arg	Asn	Lys	Asp	Val	Lys	Gln	Ala	Leu	Lys	Asn	Val	Leu	
					290				295				300			
aga	tga															918
Arg	*															
305																

<210> 7
<211> 612
<212> DNA
<213> Homo sapiens

<400> 7																	
atg	gtt	aga	gga	aat	tct	act	ttg	gtg	acg	gaa	ttt	att	ctc	ttg	gga	48	
Met	Val	Arg	Gly	Asn	Ser	Thr	Leu	Val	Thr	Glu	Phe	Ile	Leu	Leu	Gly		
1					5				10				15				
tta	aag	gat	ctt	cca	gag	ctt	cag	ccc	atc	ctc	ttt	gta	ctg	ttc	ctg	96	
Leu	Lys	Asp	Leu	Pro	Glu	Leu	Gln	Pro	Ile	Leu	Phe	Val	Leu	Phe	Leu		
					20				25				30				
cta	atc	tac	atc	act	gtc	ggg	ggg	aac	ctt	ggg	atg	ttg	gtg	ttg	ttg	144	
Leu	Ile	Tyr	Leu	Ile	Thr	Val	Gly	Gly	Asn	Leu	Gly	Met	Leu	Val	Leu		
					35				40				45				
atc	agg	ata	gat	tca	cgc	ctc	cac	acc	ccc	atg	tat	ttc	ttt	ctt	gct	192	
Ile	Arg	Ile	Asp	Ser	Arg	Leu	His	Thr	Pro	Met	Tyr	Phe	Phe	Leu	Ala		
					50				55				60				
agt	ttg	tcc	tgc	ttg	gat	ttg	tat	tac	tcc	act	aat	gtg	act	ccc	aag	240	
Ser	Leu	Ser	Cys	Leu	Asp	Leu	Tyr	Tyr	Ser	Thr	Asn	Val	Thr	Pro	Lys		
					65				70				75		80		
atg	ttg	gtg	aac	ttc	tca	gac	aag	aaa	gcc	att	tcc	tat	gct	gct	288		
Met	Leu	Val	Asn	Phe	Phe	Ser	Asp	Lys	Lys	Ala	Ile	Ser	Tyr	Ala	Ala		
					85				90				95				
tgt	tta	gtc	cag	tgc	tat	ttt	ttc	att	gct	gtg	gtg	att	act	gaa	tat	336	
Cys	Leu	Val	Gln	Cys	Tyr	Phe	Phe	Ile	Ala	Val	Val	Ile	Thr	Glu	Tyr		
					100				105				110				
tat	atg	cta	gct	gta	atg	gcc	tat	gat	agg	tat	gtg	gcc	atc	tgt	aac	384	
Tyr	Met	Leu	Ala	Val	Met	Ala	Tyr	Asp	Arg	Tyr	Val	Ala	Ile	Cys	Asn		
					115				120				125				

cct ttg ctt tac agc agc aag atg tcc aaa ggg ctc tgt att cgc ctg	432
Pro Leu Leu Tyr Ser Ser Lys Met Ser Lys Gly Leu Cys Ile Arg Leu	
130 135 140	
att gct ggt cca tat gtc tat ggg ttt ctt agt gga ctg atg gaa acc	480
Ile Ala Gly Pro Tyr Val Tyr Gly Phe Leu Ser Gly Leu Met Glu Thr	
145 150 155 160	
atg tgg aca tac cac ttg acc ttc tgt ggc tcc aat atc att aat cac	528
Met Trp Thr Tyr His Leu Thr Phe Cys Gly Ser Asn Ile Ile Asn His	
165 170 175	
ttc tac tgt gct gac cca ccc ctc atc cga ctt tcc tgc tct gac act	576
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Arg Leu Ser Cys Ser Asp Thr	
180 185 190	
ttc att aag gaa aca tcc atg ttt gtg gta gca tga	612
Phe Ile Lys Glu Thr Ser Met Phe Val Val Ala *	
195 200	

<210> 8
<211> 807
<212> DNA
<213> Homo sapiens

<400> 8	
ttg ccc tca tcc agg cca acg ccc cgg ctc cac acg ccc atg tac ttt	48
Leu Pro Ser Ser Arg Pro Thr Pro Arg Leu His Thr Pro Met Tyr Phe	
1 5 10 15	
ttc ctg agc aac tta tcc ttt gtg gat ctg tgc ttc tct tcc aat gtg	96
Phe Leu Ser Asn Leu Ser Phe Val Asp Leu Cys Phe Ser Ser Asn Val	
20 25 30	
act cca agg atg ctg gag att ttc ctt tca gag aag aaa agc att tcc	144
Thr Pro Arg Met Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser	
35 40 45	
tat cct gcc cgt ctt gtg cag tgt tac ctt ttt atc acc ttg gtc cac	192
Tyr Pro Ala Arg Leu Val Gln Cys Tyr Leu Phe Ile Thr Leu Val His	
50 55 60	
gtt gag ctc tac atc ctg gct gtg atg gcc ttt gac cgg tac atg gcc	240
Val Glu Leu Tyr Ile Leu Ala Val Met Ala Phe Asp Arg Tyr Met Ala	
65 70 75 80	
atc tgc aac cct ctg ctt tat ggc agc aga atg tcc aag agc gtg tgc	288
Ile Cys Asn Pro Leu Leu Tyr Gly Ser Arg Met Ser Lys Ser Val Cys	
85 90 95	
tct ttc ctc atc aca gtg ctt tat gtg tat gga gca ctc act ggc ctg	336
Ser Phe Leu Ile Thr Val Leu Tyr Val Tyr Gly Ala Leu Thr Gly Leu	
100 105 110	
atg gag act atg tgg acc tac aac cta gcc ttc tgt ggc ccc agt gaa	384
Met Glu Thr Met Trp Thr Tyr Asn Leu Ala Phe Cys Gly Pro Ser Glu	
115 120 125	
att aat cac ttc tac tgt gtg gac cca cca ctg att aag ctg gct tgt	432
Ile Asn His Phe Tyr Cys Val Asp Pro Pro Leu Ile Lys Leu Ala Cys	
130 135 140	
tct gac acc tac aac aag gag gtg tca atg ttt gtt gtg gct ggt ttc	480
Ser Asp Thr Tyr Asn Lys Glu Val Ser Met Phe Val Val Ala Gly Phe	
145 150 155 160	
aac ttc act tat cct ctc ctt atc atc ctc att tcc tat ctc tac ata	528
Asn Phe Thr Tyr Pro Leu Leu Ile Ile Leu Ile Ser Tyr Leu Tyr Ile	
165 170 175	
ttt cct gcc acc cta agg atc tgc tct aca gaa ggc agg cac aaa gct	576
Phe Pro Ala Thr Leu Arg Ile Cys Ser Thr Glu Gly Arg His Lys Ala	
180 185 190	
ttt tct acc tgt ggc tcc cat ctg aca gcc gtt act att ttc tat tca	624
Phe Ser Thr Cys Gly Ser His Leu Thr Ala Val Thr Ile Phe Tyr Ser	
195 200 205	
gct ctt ttc ttc atg tat ctc aga cgt cca tca gaa gag tcc atg gag	672
Ala Leu Phe Phe Met Tyr Leu Arg Arg Pro Ser Glu Glu Ser Met Glu	

52

210	215	220															
cag	ggg	aaa	atg	gta	gct	gta	ttt	tat	acc	act	gta	atc	ccc	atg	ttg		720
Gln	Gly	Lys	Met	Val	Ala	Val	Phe	Tyr	Thr	Thr	Val	Ile	Pro	Met	Leu		
225				230					235						240		
aat	ccc	atg	atc	tac	agt	ctg	agg	aac	aaa	gat	gtg	aaa	gag	gca	tta		768
Asn	Pro	Met	Ile	Tyr	Ser	Leu	Arg	Asn	Lys	Asp	Val	Lys	Glu	Ala	Leu		
															245		
tgc	aaa	gaa	ctg	ttc	aaa	aga	aaa	ttg	ttt	tct	aaa	taa					807
Cys	Lys	Glu	Leu	Phe	Lys	Arg	Lys	Leu	Phe	Ser	Lys	*					
															260		
															265		

<210> 9
<211> 363
<212> DNA
<213> Homo sapiens

<400> 9																	
atg	aga	agg	aac	ttc	acg	ttg	gtg	act	gag	ttc	att	ctc	ctg	gga	ctg		48
Met	Arg	Arg	Asn	Phe	Thr	Leu	Val	Thr	Glu	Phe	Ile	Leu	Leu	Gly	Leu		
1					5				10					15			
acg	aat	cac	cag	gaa	tta	cag	att	ctc	ctc	ttc	atg	ctg	ttt	ctg	gcc		96
Thr	Asn	His	Gln	Glu	Leu	Gln	Ile	Leu	Leu	Phe	Met	Leu	Phe	Leu	Ala		
							20		25				30				
att	tac	atg	gtc	aca	gtg	gca	ggg	aat	ctt	agc	atg	att	gcc	ctc	atc		144
Ile	Tyr	Met	Val	Thr	Val	Ala	Gly	Asn	Leu	Ser	Met	Ile	Ala	Leu	Ile		
					35			40			45						
cag	gcc	aat	gcc	cgg	ctc	cac	acg	ccc	atg	tac	ttt	ttc	ctg	agc	cac		192
Gln	Ala	Asn	Ala	Arg	Leu	His	Thr	Pro	Met	Tyr	Phe	Leu	Ser	His			
					50		55		60								
tta	tcc	ttc	ctg	gat	ctg	tgc	ttc	tct	tcc	aat	gtg	acc	cca	aag	atg		240
Leu	Ser	Phe	Leu	Asp	Leu	Cys	Phe	Ser	Asn	Val	Thr	Pro	Lys	Met			
					65		70		75			80					
ctg	gag	att	ttc	ctt	tca	gag	aag	aaa	agc	att	tcc	tat	cct	gcc	tgt		288
Leu	Glu	Ile	Phe	Leu	Ser	Glu	Lys	Lys	Ser	Ile	Ser	Tyr	Pro	Ala	Cys		
					85		90		95								
ctt	gtt	cag	tgt	tac	ctt	tat	atc	atc	ttg	gta	cac	gtt	gag	atc	tac		336
Leu	Val	Gln	Cys	Tyr	Leu	Tyr	Ile	Ile	Leu	Val	His	Val	Glu	Ile	Tyr		
					100		105		110								
atc	ctg	gct	gtg	atg	gcc	ttt	gac	tag									363
Ile	Leu	Ala	Val	Met	Ala	Phe	Asp	*									
					115		120										

<210> 10
<211> 936
<212> DNA
<213> Homo sapiens

<400> 10																	
atg	aga	aga	aac	tgc	acg	ttg	gtg	act	gag	ttc	att	ctc	ctg	gga	ctg		48
Met	Arg	Arg	Asn	Cys	Thr	Leu	Val	Thr	Glu	Phe	Ile	Leu	Leu	Gly	Leu		
1					5				10					15			
acc	agt	cgc	cgg	gaa	tta	caa	att	ctc	ctc	ttc	acg	ctg	ttt	ctg	gcc		96
Thr	Ser	Arg	Arg	Glu	Leu	Gln	Ile	Leu	Phe	Thr	Leu	Phe	Leu	Ala			
					20		25		30								
att	tac	atg	gtc	acg	gtg	gca	ggg	aac	ctt	ggc	atg	att	gtc	ctc	atc		144
Ile	Tyr	Met	Val	Thr	Val	Ala	Gly	Asn	Leu	Gly	Met	Ile	Val	Leu	Ile		
					35		40		45								
cag	gcc	aac	gcc	tgg	ctc	cac	atg	tac	ttt	ttc	ctg	agc	cac				192
Gln	Ala	Asn	Ala	Trp	Leu	His	Met	Pro	Met	Tyr	Phe	Phe	Leu	Ser	His		
					50		55		60								
tta	tcc	ttc	gtg	gat	ctg	tgc	ttc	tct	tcc	aat	gtg	act	cca	aag	atg		240
Leu	Ser	Phe	Val	Asp	Leu	Cys	Phe	Ser	Ser	Asn	Val	Thr	Pro	Lys	Met		
					65		70		75			80					

ctg gag att ttc ctt tca gag aag aaa agc att tcc tat cct gcc tgt	288
Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser Tyr Pro Ala Cys	
85 90 95	
ctt gtg cag tgt tac ctt ttt atc gcc ttg gtc cat gtt gag atc tac	336
Leu Val Gln Cys Tyr Leu Phe Ile Ala Leu Val His Val Glu Ile Tyr	
100 105 110	
atc ctg gct gtg atg gcc ttt gac cgg tac atg gcc atc tgc aac cct	384
Ile Leu Ala Val Met Ala Phe Asp Arg Tyr Met Ala Ile Cys Asn Pro	
115 120 125	
ctg ctt tat ggc agc aga atg tcc aag agt gtg tgc tcc ttc ctc atc	432
Leu Leu Tyr Gly Ser Arg Met Ser Lys Ser Val Cys Ser Phe Leu Ile	
130 135 140	
acg gtg cct tat gtg tat gga gcg ctc act ggc ctg atg gag acc atg	480
Thr Val Pro Tyr Val Tyr Gly Ala Leu Thr Gly Leu Met Glu Thr Met	
145 150 155 160	
tgg acc tac aac cta gcc ttc tgt ggc ccc aat gaa att aat cac ttc	528
Trp Thr Tyr Asn Leu Ala Phe Cys Gly Pro Asn Glu Ile Asn His Phe	
165 170 175	
tac tgt gcg gac cca cca ctg att aag ctg gct tgt tct gac acc tac	576
Tyr Cys Ala Asp Pro Pro Leu Ile Lys Leu Ala Cys Ser Asp Thr Tyr	
180 185 190	
aac aag gag ttg tca atg ttt att gtg gct ggc tgg aac ctt tct ttt	624
Asn Lys Glu Leu Ser Met Phe Ile Val Ala Gly Trp Asn Leu Ser Phe	
195 200 205	
tct ctc ttc atc ata tgt att tcc tac ctt tac att ttc cct gct att	672
Ser Leu Phe Ile Ile Cys Ile Ser Tyr Leu Tyr Ile Phe Pro Ala Ile	
210 215 220	
tta aag att cgc tct aca gag ggc agg caa aaa gct ttt tct acc tgt	720
Leu Lys Ile Arg Ser Thr Glu Gly Arg Gln Lys Ala Phe Ser Thr Cys	
225 230 235 240	
ggc tcc cat ctg aca gct gtc act ata ttc tat gca acc ctt ttc ttc	768
Gly Ser His Leu Thr Ala Val Thr Ile Phe Tyr Ala Thr Leu Phe Phe	
245 250 255	
atg tat ctc aga ccc ccc tca aag gaa tct gtt gaa cag ggt aaa atg	816
Met Tyr Leu Arg Pro Pro Ser Lys Glu Ser Val Glu Gln Gly Lys Met	
260 265 270	
gta gct gta ttt tat acc aca gta atc cct atg ctg aac ctt ata att	864
Val Ala Val Phe Tyr Thr Val Ile Pro Met Leu Asn Leu Ile Ile	
275 280 285	
tat agc ctt aga aat aaa aat gta aaa gaa gca tta atc aaa gag ctg	912
Tyr Ser Leu Arg Asn Lys Asn Val Lys Glu Ala Leu Ile Lys Glu Leu	
290 295 300	
tca atg aag ata tac ttt tct taa	936
Ser Met Lys Ile Tyr Phe Ser *	
305 310	

<210> 11
<211> 180
<212> DNA
<213> Homo sapiens

<400> 11	
atg tcc aga aga aac tat act gaa ctg aca gaa ttt gtt ctc ttg ggt	48
Met Ser Arg Arg Asn Tyr Thr Glu Leu Thr Glu Phe Val Leu Leu Gly	
1 5 10 15	
cta aca agc cgt cca gag ctg cga gtt gct ttc ttg gca ctg ttc ctt	96
Leu Thr Ser Arg Pro Glu Leu Arg Val Ala Phe Leu Ala Leu Phe Leu	
20 25 30	
ttt gtc tac ata gcc act gtg gta gga aac ttg ggg atg att att tta	144
Phe Val Tyr Ile Ala Thr Val Val Gly Asn Leu Gly Met Ile Ile Leu	
35 40 45	
atc aaa gtt gat tct cga ctt cac act ccc atg taa	180
Ile Lys Val Asp Ser Arg Leu His Thr Pro Met *	

50	55	54	60												
<210> 12															
<211> 64															
<212> PRT															
<213> Homo sapiens															
<400> 12															
Met	Ser	Phe	Leu	Ile	Arg	Ser	Asp	Ser	Thr	Leu	His	Thr	Pro	Met	Cys
1				5					10					15	
Leu	Phe	Leu	Ser	His	Leu	Ser	Phe	Val	Asp	Leu	Tyr	Tyr	Ala	Thr	Asn
				20				25					30		
Ala	Thr	Pro	Pro	Met	Leu	Val	Asn	Phe	Phe	Phe	Pro	Arg	Glu	Lys	Pro
				35			40				45				
Phe	Pro	Leu	Leu	Val	Ala	Leu	Ser	Asn	Phe	Thr	Phe	Ser	Leu	His	Trp
					50						60				
<210> 13															
<211> 315															
<212> PRT															
<213> Homo sapiens															
<400> 13															
Met	Phe	Ser	Pro	Asn	His	Thr	Ile	Val	Thr	Glu	Phe	Ile	Leu	Leu	Gly
1					5				10				15		
Leu	Thr	Asp	Asp	Pro	Val	Leu	Glu	Lys	Ile	Leu	Phe	Gly	Val	Phe	Leu
					20			25				30			
Ala	Ile	Tyr	Leu	Ile	Thr	Leu	Ala	Gly	Asn	Leu	Cys	Met	Ile	Leu	Leu
					35			40			45				
Ile	Arg	Thr	Asn	Ser	His	Leu	Gln	Thr	Pro	Met	Tyr	Phe	Phe	Leu	Gly
					50		55			60					
His	Leu	Ser	Phe	Val	Asp	Ile	Cys	Tyr	Ser	Ser	Asn	Val	Thr	Pro	Asn
					65		70			75			80		
Met	Leu	His	Asn	Phe	Leu	Ser	Glu	Gln	Lys	Thr	Ile	Ser	Tyr	Ala	Gly
					85			90			95				
Cys	Phe	Thr	Gln	Cys	Leu	Leu	Phe	Ile	Ala	Leu	Val	Ile	Thr	Glu	Phe
					100			105			110				
Tyr	Phe	Leu	Ala	Ser	Met	Ala	Leu	Asp	Arg	Tyr	Val	Ala	Ile	Cys	Ser
					115			120			125				
Pro	Leu	His	Tyr	Ser	Ser	Arg	Met	Ser	Lys	Asn	Ile	Cys	Ile	Ser	Leu
					130		135			140					
Val	Thr	Val	Pro	Tyr	Met	Tyr	Gly	Phe	Leu	Asn	Gly	Leu	Ser	Gln	Thr
					145		150			155			160		
Leu	Leu	Thr	Phe	His	Leu	Ser	Phe	Cys	Gly	Ser	Leu	Glu	Ile	Asn	His
					165			170			175				
Phe	Tyr	Cys	Ala	Asp	Pro	Pro	Leu	Ile	Met	Leu	Ala	Cys	Ser	Asp	Thr
					180			185			190				
Arg	Val	Lys	Lys	Met	Ala	Met	Phe	Val	Val	Ala	Gly	Phe	Thr	Leu	Ser
					195		200			205					
Ser	Ser	Leu	Phe	Ile	Ile	Leu	Leu	Ser	Tyr	Leu	Phe	Ile	Phe	Ala	Ala
					210		215			220					
Ile	Phe	Arg	Ile	Arg	Ser	Ala	Glu	Gly	Arg	His	Lys	Ala	Phe	Ser	Thr
					225		230			235			240		
Cys	Ala	Ser	His	Leu	Thr	Ile	Val	Thr	Leu	Phe	Tyr	Gly	Thr	Leu	Phe
					245			250			255				
Cys	Met	Tyr	Val	Arg	Pro	Pro	Ser	Glu	Lys	Ser	Val	Glu	Glu	Ser	Lys
					260			265			270				
Ile	Thr	Ala	Val	Phe	Tyr	Thr	Phe	Leu	Thr	Pro	Met	Leu	Asn	Pro	Leu
					275		280			285					
Ile	Tyr	Ser	Leu	Arg	Asn	Thr	Asp	Val	Ile	Leu	Ala	Met	Gln	Gln	Met
					290		295			300					
Ile	Arg	Gly	Lys	Ser	Phe	His	Lys	Ile	Ala	Val					
					305		310			315					

<210> 14
<211> 172
<212> PRT
<213> Homo sapiens

<400> 14
Met Leu Lys Lys Asn His Thr Ala Val Thr Glu Phe Val Leu Leu Gly
1 5 10 15
Leu Thr Asp Arg Ala Glu Leu Gln Ser Leu Leu Phe Val Val Phe Leu
20 25 30
Val Ile Tyr Leu Ile Thr Val Ile Gly Asn Val Ser Met Ile Leu Leu
35 40 45
Ile Arg Ser Asp Ser Thr Leu His Thr Pro Met Tyr Phe Phe Leu Ser
50 55 60
His Leu Ser Phe Val Asp Leu Cys Tyr Thr Thr Asn Val Thr Pro Gln
65 70 75 80
Met Leu Val Asn Phe Leu Ser Lys Arg Lys Thr Ile Ser Phe Ile Gly
85 90 95
Cys Phe Ile Gln Phe His Phe Phe Ile Ala Leu Val Ile Thr Asp Tyr
100 105 110
Tyr Met Leu Thr Val Met Ala Tyr Asp Arg Tyr Met Ala Ile Cys Lys
115 120 125
Pro Leu Leu Tyr Gly Ser Lys Met Thr Arg Cys Val Cys Leu Cys Leu
130 135 140
Ala Ala Ala Pro Tyr Ile Tyr Gly Phe Ala Asn Gly Leu Ser Thr Asp
145 150 155 160
His Pro Asp Ala Ser Ser Val Leu Leu Trp Thr Gln
165 170

<210> 15
<211> 315
<212> PRT
<213> Homo sapiens

<400> 15
Met Leu Ser Pro Asn His Thr Ile Val Thr Glu Phe Ile Leu Leu Gly
1 5 10 15
Leu Thr Asp Asp Pro Val Leu Glu Lys Ile Leu Phe Gly Val Phe Leu
20 25 30
Ala Ile Tyr Leu Ile Thr Leu Ala Gly Asn Leu Cys Met Ile Leu Leu
35 40 45
Ile Arg Thr Asn Ser Gln Leu Gln Thr Pro Met Tyr Phe Phe Leu Gly
50 55 60
His Leu Ser Phe Leu Asp Ile Cys Tyr Ser Ser Asn Val Thr Pro Asn
65 70 75 80
Met Leu His Asn Phe Leu Ser Glu Gln Lys Thr Ile Ser Tyr Ala Gly
85 90 95
Cys Phe Thr Gln Cys Leu Leu Phe Ile Ala Leu Val Ile Thr Glu Phe
100 105 110
Tyr Phe Leu Ala Ser Met Ala Leu Asp Arg Tyr Val Ala Ile Cys Ser
115 120 125
Pro Leu His Tyr Ser Ser Arg Met Ser Lys Asn Ile Cys Ile Ser Leu
130 135 140
Val Thr Val Pro Tyr Met Tyr Gly Phe Leu Asn Gly Leu Ser Gln Thr
145 150 155 160
Leu Leu Thr Phe His Leu Ser Phe Cys Gly Ser Leu Glu Ile Asn His
165 170 175
Phe Tyr Cys Ala Asp Pro Pro Leu Ile Met Leu Ala Cys Ser Asp Thr
180 185 190
Arg Val Lys Lys Met Ala Met Phe Val Val Ala Gly Phe Thr Leu Ser
195 200 205
Ser Ser Leu Phe Ile Ile Leu Leu Ser Tyr Leu Phe Ile Phe Ala Ala

210	215	220
Ile Phe Arg Ile Arg Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr		
225 230	235	240
Cys Ala Ser His Leu Thr Ile Val Thr Leu Phe Tyr Gly Thr Leu Phe		
245 250	255	
Cys Met Tyr Val Arg Pro Pro Ser Glu Lys Ser Val Glu Glu Ser Lys		
260 265	270	
Ile Ile Ala Val Phe Tyr Thr Phe Leu Ser Pro Met Leu Asn Pro Leu		
275 280	285	
Ile Tyr Ser Leu Arg Asn Arg Asp Val Ile Leu Ala Ile Gln Gln Met		
290 295	300	
Ile Arg Gly Lys Ser Phe Cys Lys Ile Ala Val		
305 310	315	

<210> 16

<211> 305

<212> PRT

<213> Homo sapiens

<400> 16

Met Ser Asn Thr Asn Gly Ser Ala Ile Thr Glu Phe Ile Leu Leu Gly		
1 5	10	15
Leu Thr Asp Cys Pro Glu Leu Gln Ser Leu Leu Phe Val Leu Phe Leu		
20 25	30	
Val Val Tyr Leu Val Thr Leu Leu Gly Asn Leu Gly Met Ile Met Leu		
35 40	45	
Met Arg Leu Asp Ser Arg Leu His Thr Pro Met Tyr Phe Phe Leu Thr		
50 55	60	
Asn Leu Ala Phe Val Asp Leu Cys Tyr Thr Ser Asn Ala Thr Pro Gln		
65 70	75	80
Met Ser Thr Asn Ile Val Ser Glu Lys Thr Ile Ser Phe Ala Gly Cys		
85 90	95	
Phe Thr Gln Cys Tyr Ile Phe Ile Ala Leu Leu Thr Glu Phe Tyr		
100 105	110	
Met Leu Ala Ala Met Ala Tyr Asp Arg Tyr Val Ala Ile Tyr Asp Pro		
115 120	125	
Leu Arg Tyr Ser Val Lys Thr Ser Arg Arg Val Cys Ile Cys Leu Ala		
130 135	140	
Thr Phe Pro Tyr Val Tyr Gly Phe Ser Asp Gly Leu Phe Gln Ala Ile		
145 150	155	160
Leu Thr Phe Arg Leu Thr Phe Cys Arg Ser Asn Val Ile Asn His Phe		
165 170	175	
Tyr Cys Ala Asp Pro Pro Leu Ile Lys Leu Ser Cys Ser Asp Thr Tyr		
180 185	190	
Val Lys Glu His Ala Met Phe Ile Ser Ala Gly Phe Asn Leu Ser Ser		
195 200	205	
Ser Leu Thr Ile Val Leu Val Ser Tyr Ala Phe Ile Leu Ala Ala Ile		
210 215	220	
Leu Arg Ile Lys Ser Ala Glu Gly Arg His Lys Ala Phe Ser Thr Cys		
225 230	235	240
Gly Ser His Met Met Ala Val Thr Leu Phe Tyr Gly Thr Leu Phe Cys		
245 250	255	
Met Tyr Ile Arg Pro Pro Thr Asp Lys Thr Val Glu Glu Ser Lys Ile		
260 265	270	
Ile Ala Val Phe Tyr Thr Phe Val Ser Pro Val Leu Asn Pro Leu Ile		
275 280	285	
Tyr Ser Leu Arg Asn Lys Asp Val Lys Gln Ala Leu Lys Asn Val Leu		
290 295	300	

Arg

305

<210> 17

<211> 203

<212> PRT

<213> Homo sapiens

<400> 17

Met	Val	Arg	Gly	Asn	Ser	Thr	Leu	Val	Thr	Glu	Phe	Ile	Leu	Leu	Gly
1							5		10						15
Leu	Lys	Asp	Leu	Pro	Glu	Leu	Gln	Pro	Ile	Leu	Phe	Val	Leu	Phe	Leu
							20		25						30
Leu	Ile	Tyr	Leu	Ile	Thr	Val	Gly	Gly	Asn	Leu	Gly	Met	Leu	Val	Leu
							35		40						45
Ile	Arg	Ile	Asp	Ser	Arg	Leu	His	Thr	Pro	Met	Tyr	Phe	Phe	Leu	Ala
							50		55						60
Ser	Leu	Ser	Cys	Leu	Asp	Leu	Tyr	Tyr	Ser	Thr	Asn	Val	Thr	Pro	Lys
							65		70						80
Met	Leu	Val	Asn	Phe	Phe	Ser	Asp	Lys	Lys	Ala	Ile	Ser	Tyr	Ala	Ala
							85		90						95
Cys	Leu	Val	Gln	Cys	Tyr	Phe	Phe	Ile	Ala	Val	Val	Ile	Thr	Glu	Tyr
							100		105						110
Tyr	Met	Leu	Ala	Val	Met	Ala	Tyr	Asp	Arg	Tyr	Val	Ala	Ile	Cys	Asn
							115		120						125
Pro	Leu	Leu	Tyr	Ser	Ser	Lys	Met	Ser	Lys	Gly	Leu	Cys	Ile	Arg	Leu
							130		135						140
Ile	Ala	Gly	Pro	Tyr	Val	Tyr	Gly	Phe	Leu	Ser	Gly	Leu	Met	Glu	Thr
							145		150						160
Met	Trp	Thr	Tyr	His	Leu	Thr	Phe	Cys	Gly	Ser	Asn	Ile	Ile	Asn	His
							165		170						175
Phe	Tyr	Cys	Ala	Asp	Pro	Pro	Leu	Ile	Arg	Leu	Ser	Cys	Ser	Asp	Thr
							180		185						190
Phe	Ile	Lys	Glu	Thr	Ser	Met	Phe	Val	Val	Ala					
							195		200						

<210> 18

<211> 268

<212> PRT

<213> Homo sapiens

<400> 18

Leu	Pro	Ser	Ser	Arg	Pro	Thr	Pro	Arg	Leu	His	Thr	Pro	Met	Tyr	Phe
1							5		10						15
Phe	Leu	Ser	Asn	Leu	Ser	Phe	Val	Asp	Leu	Cys	Phe	Ser	Ser	Asn	Val
							20		25						30
Thr	Pro	Arg	Met	Leu	Glu	Ile	Phe	Leu	Ser	Glu	Lys	Lys	Ser	Ile	Ser
							35		40						45
Tyr	Pro	Ala	Arg	Leu	Val	Gln	Cys	Tyr	Leu	Phe	Ile	Thr	Leu	Val	His
							50		55						60
Val	Glu	Leu	Tyr	Ile	Leu	Ala	Val	Met	Ala	Phe	Asp	Arg	Tyr	Met	Ala
							65		70						80
Ile	Cys	Asn	Pro	Leu	Leu	Tyr	Gly	Ser	Arg	Met	Ser	Lys	Ser	Val	Cys
							85		90						95
Ser	Phe	Leu	Ile	Thr	Val	Leu	Tyr	Val	Tyr	Gly	Ala	Leu	Thr	Gly	Leu
							100		105						110
Met	Glu	Thr	Met	Trp	Thr	Tyr	Asn	Leu	Ala	Phe	Cys	Gly	Pro	Ser	Glu
							115		120						125
Ile	Asn	His	Phe	Tyr	Cys	Val	Asp	Pro	Pro	Leu	Ile	Lys	Leu	Ala	Cys
							130		135						140
Ser	Asp	Thr	Tyr	Asn	Lys	Glu	Val	Ser	Met	Phe	Val	Val	Ala	Gly	Phe
							145		150						160
Asn	Phe	Thr	Tyr	Pro	Leu	Leu	Ile	Ile	Leu	Ile	Ser	Tyr	Leu	Tyr	Ile
							165		170						175
Phe	Pro	Ala	Thr	Leu	Arg	Ile	Cys	Ser	Thr	Glu	Gly	Arg	His	Lys	Ala
							180		185						190
Phe	Ser	Thr	Cys	Gly	Ser	His	Leu	Thr	Ala	Val	Thr	Ile	Phe	Tyr	Ser
							195		200						205

Ala Leu Phe Phe Met Tyr Leu Arg Arg Pro Ser Glu Glu Ser Met Glu
 210 215 220
 Gln Gly Lys Met Val Ala Val Phe Tyr Thr Thr Val Ile Pro Met Leu
 225 230 235 240
 Asn Pro Met Ile Tyr Ser Leu Arg Asn Lys Asp Val Lys Glu Ala Leu
 245 250 255
 Cys Lys Glu Leu Phe Lys Arg Lys Leu Phe Ser Lys
 260 265

<210> 19
 <211> 120
 <212> PRT
 <213> Homo sapiens

<400> 19
 Met Arg Arg Asn Phe Thr Leu Val Thr Glu Phe Ile Leu Leu Gly Leu
 1 5 10 15
 Thr Asn His Gln Glu Leu Gln Ile Leu Leu Phe Met Leu Phe Leu Ala
 20 25 30
 Ile Tyr Met Val Thr Val Ala Gly Asn Leu Ser Met Ile Ala Leu Ile
 35 40 45
 Gln Ala Asn Ala Arg Leu His Thr Pro Met Tyr Phe Phe Leu Ser His
 50 55 60
 Leu Ser Phe Leu Asp Leu Cys Phe Ser Ser Asn Val Thr Pro Lys Met
 65 70 75 80
 Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser Tyr Pro Ala Cys
 85 90 95
 Leu Val Gln Cys Tyr Leu Tyr Ile Ile Leu Val His Val Glu Ile Tyr
 100 105 110
 Ile Leu Ala Val Met Ala Phe Asp
 115 120

<210> 20
 <211> 311
 <212> PRT
 <213> Homo sapiens

<400> 20
 Met Arg Arg Asn Cys Thr Leu Val Thr Glu Phe Ile Leu Leu Gly Leu
 1 5 10 15
 Thr Ser Arg Arg Glu Leu Gln Ile Leu Leu Phe Thr Leu Phe Leu Ala
 20 25 30
 Ile Tyr Met Val Thr Val Ala Gly Asn Leu Gly Met Ile Val Leu Ile
 35 40 45
 Gln Ala Asn Ala Trp Leu His Met Pro Met Tyr Phe Phe Leu Ser His
 50 55 60
 Leu Ser Phe Val Asp Leu Cys Phe Ser Ser Asn Val Thr Pro Lys Met
 65 70 75 80
 Leu Glu Ile Phe Leu Ser Glu Lys Lys Ser Ile Ser Tyr Pro Ala Cys
 85 90 95
 Leu Val Gln Cys Tyr Leu Phe Ile Ala Leu Val His Val Glu Ile Tyr
 100 105 110
 Ile Leu Ala Val Met Ala Phe Asp Arg Tyr Met Ala Ile Cys Asn Pro
 115 120 125
 Leu Leu Tyr Gly Ser Arg Met Ser Lys Ser Val Cys Ser Phe Leu Ile
 130 135 140
 Thr Val Pro Tyr Val Tyr Gly Ala Leu Thr Gly Leu Met Glu Thr Met
 145 150 155 160
 Trp Thr Tyr Asn Leu Ala Phe Cys Gly Pro Asn Glu Ile Asn His Phe
 165 170 175
 Tyr Cys Ala Asp Pro Pro Leu Ile Lys Leu Ala Cys Ser Asp Thr Tyr
 180 185 190
 Asn Lys Glu Leu Ser Met Phe Ile Val Ala Gly Trp Asn Leu Ser Phe

195	200	205	
Ser Leu Phe Ile Ile Cys Ile	Ser Tyr Leu Tyr Ile	Phe Pro Ala Ile	
210	215	220	
Leu Lys Ile Arg Ser Thr Glu Gly Arg Gln Lys	Ala Phe Ser Thr Cys		
225	230	235	240
Gly Ser His Leu Thr Ala Val Thr Ile Phe Tyr	Ala Thr Leu Phe Phe		
245	250	255	
Met Tyr Leu Arg Pro Pro Ser Lys Glu Ser Val	Glu Gln Gly Lys Met		
260	265	270	
Val Ala Val Phe Tyr Thr Thr Val Ile Pro Met	Leu Asn Leu Ile Ile		
275	280	285	
Tyr Ser Leu Arg Asn Lys Asn Val Lys Glu Ala	Leu Ile Lys Glu Leu		
290	295	300	
Ser Met Lys Ile Tyr Phe Ser			
305	310		

<210> 21

<211> 59

<212> PRT

<213> Homo sapiens

<400> 21

Met Ser Arg Arg Asn Tyr Thr Glu Leu Thr Glu Phe Val	Leu Leu Gly		
1	5	10	15
Leu Thr Ser Arg Pro Glu Leu Arg Val Ala Phe Leu Ala	Leu Phe Leu		
20	25	30	
Phe Val Tyr Ile Ala Thr Val Val Gly Asn Leu Gly	Met Ile Ile Leu		
35	40	45	
Ile Lys Val Asp Ser Arg Leu His Thr Pro Met			
50	55		

<210> 22

<211> 30

<212> DNA

<213> Artificial Sequence

<400> 22

cctggagggt ttcaaaggct gataactttag

30

<210> 23

<211> 26

<212> DNA

<213> Artificial Sequence

<400> 23

ctccagcctg agcaacagag caatac

26

<210> 24

<211> 30

<212> DNA

<213> Artificial Sequence

<400> 24

ctcacattca ttgttcttca cagaccgac

30

<210> 25

<211> 24

<212> DNA

<213> Artificial Sequence

<400> 25

ccctgtggg atctggatca agac

24

<210> 26
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerPU

<400> 26
tgtaaaacga cggccagt

18

<210> 27
<211> 18
<212> DNA
<213> Artificial Sequence

<220>
<223> sequencing oligonucleotide PrimerRP

<400> 27
cagggaaacag ctatgacc

18